US006438523B1

(12) **United States Patent**
**Oberteuffer et al.**

(10) **Patent No.: US 6,438,523 B1**
(45) **Date of Patent: Aug. 20, 2002**

(54) **PROCESSING HANDWRITTEN AND HAND-DRAWN INPUT AND SPEECH INPUT**

(76) Inventors: **John A. Oberteuffer**, 14 Glen Rd. S., Lexington, MA (US) 02420; **John Wilbanks**, 78 Porter Rd. #22, Cambridge, MA (US) 02140; **Kyung-Ho Loken-Kim**, 31 Robbins Rd., Lexington, MA (US) 02421; **William Kania**, 21 Wayside Rd., Westborough, MA (US) 01581

( * ) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **09/313,664**

(22) Filed: **May 18, 1999**

**Related U.S. Application Data**

(60) Provisional application No. 60/086,346, filed on May 20, 1998.

(51) **Int. Cl.**[7] ..... **G10L 21/00**; G06K 9/03

(52) **U.S. Cl.** ..... **704/270**; 704/251; 382/186; 382/187

(58) **Field of Search** ..... 704/235, 236, 704/255, 270, 260, 251; 382/187, 186, 189; 345/157

(56) **References Cited**

U.S. PATENT DOCUMENTS

| | | | |
|---|---|---|---|
| 4,401,855 | A | 8/1983 | Broderson et al. |
| 4,689,761 | A | 8/1987 | Yurchenco |
| 4,748,674 | A | 5/1988 | Freeman |
| 5,157,384 | A | 10/1992 | Greanias et al. |
| 5,177,685 | A | 1/1993 | Davis et al. |
| 5,208,786 | A | 5/1993 | Weinstein et al. |
| 5,377,303 | A | 12/1994 | Firman |
| 5,404,524 | A | 4/1995 | Celi, Jr. |
| 5,459,798 | A | 10/1995 | Bailey et al. |
| 5,463,696 | A | 10/1995 | Beernink et al. |
| 5,502,774 | A | 3/1996 | Bellegarda et al. |
| 5,528,726 | A | 6/1996 | Cook |
| 5,546,538 | A | 8/1996 | Cobbley et al. |
| 5,546,565 | A * | 8/1996 | Suzuki ..... 382/187 |
| 5,583,542 | A | 12/1996 | Capps et al. |
| 5,583,946 | A | 12/1996 | Courdol |
| 5,592,589 | A | 1/1997 | Poon |
| 5,600,765 | A | 2/1997 | Ando et al. |
| 5,606,702 | A | 2/1997 | Diel et al. |
| 5,615,299 | A | 3/1997 | Bahl et al. |
| 5,621,809 | A | 4/1997 | Bellegarda et al. |
| 5,621,857 | A | 4/1997 | Cole et al. |
| 5,625,749 | A | 4/1997 | Goldenthal et al. |

(List continued on next page.)

FOREIGN PATENT DOCUMENTS

| | | |
|---|---|---|
| JP | 61240361 | 10/1986 |
| JP | 06006436 | 1/1994 |
| JP | 6-131108 | 5/1994 |
| JP | 06131108 | 5/1994 |
| JP | 08180022 | 7/1996 |
| JP | 10083195 | 3/1998 |

OTHER PUBLICATIONS

Luc Julia and Adam Cheyer, "A Multimodal Computer–augmented interface for Distributed Applications"; Symbiosis of Human and Artifact; Elsevier 1995, pp. 237–240.

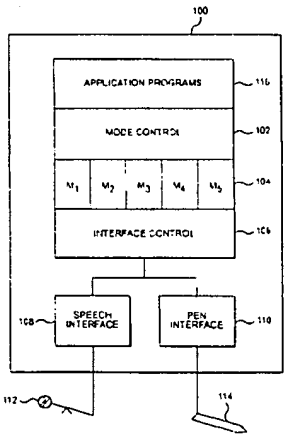
(List continued on next page.)

*Primary Examiner*—David D. Knepper
(74) *Attorney, Agent, or Firm*—Finnegan, Henderson, Farabow, Garrett & Dunner, L.L.P.

(57) **ABSTRACT**

Techniques consistent with this invention process handwritten or hand-drawn input and speech input. Method steps include recognizing received handwritten or hand-drawn input, recognizing received speech input, and creating or modifying an electronic document according to the speech or handwritten or hand-drawn input. An apparatus includes structure for recognizing handwritten or hand-drawn input, structure for recognizing speech input, and structure for activating modes for processing the handwritten or hand-drawn input and the speech input responsive to commands in the handwritten or hand-drawn input and the speech input.

**40 Claims, 10 Drawing Sheets**



U.S. PATENT DOCUMENTS

| | | | | |
|---|---|---|---|---|
| 5,633,954 | A | | 5/1997 | Gupta et al. |
| 5,655,148 | A | | 8/1997 | Richman et al. |
| 5,666,139 | A | | 9/1997 | Thielens et al. |
| 5,668,573 | A | | 9/1997 | Favot et al. |
| 5,687,221 | A | | 11/1997 | Oheda et al. |
| 5,701,393 | A | | 12/1997 | Smith et al. |
| 5,712,957 | A | | 1/1998 | Waibel et al. |
| 5,781,179 | A | * | 7/1998 | Nakajima et al. ..... 704/251 |
| 5,855,000 | A | * | 12/1998 | Waibel et al. ..... 704/235 |

OTHER PUBLICATIONS

Minh Tue Vo and Cindy Wood; "Building an Application Framework for Speech and Pen Input Integration in Multimodal Learing Interfaces"; 1996 IEEE; pp. 3545–3548.

Luc Julia and Claudie Faure; "Pattern Recognition and Beautification for Pen Based Interface"; 1995 IEEE Comput. Soc. Press; Proceedings of the Third International Conference on Document Analysis and Recognition; vol. 1; pp. 58–63.

Sharon L. Oviatt, Philip R. Cohen, and Michelle Wang; Toward Interface Design for Human Language Technology: Modality and Structure as Determinants of Linguistic Complexity; 1994 Elsevier Science B.V.; Speech Communication 1994; pp. 283–300; vol. 15, Nos. 3–4.

Bernhard Suhm; "Empirical Evaluation of Interactive Multimodal Error Connection"; 1997 IEEE Workshop on Automatic Speech Recognition and Understanding Proceedings; pp. 583–590.

Fonix Press Release; Fonix Announces Pen/Voice™ Solution for Public Safety; http://www.fonix.com/fonixnews/pressreleases/1999/feb/press021999.htm; printed Apr. 4, 2000; pp. 1–3.

Sharon Oviatt and Erik Olsen; "Integration Themes in Multimodal Human—Computer Interaction"; 1994 International Conference on Spoken Language Processing Sep. 18–22, 1994; pp. 551–554.

Adam Cheyer and Luc Julia; "Multimodal Maps: An Agent-Based Approach"; Multimodal Human—Computer Communications; pp. 111–121; Springer; 1998.

Sharon L. Oviatt, Philip R. Cohen, and Michelle Wang; "Reducing Linguistic Variability in Speech and Handwriting through Selection of Presentation Format"; Computer Dialogue Laboratory & Artificial Intelligence Center SRI International, Menlo Park, California.

International Search Report dated Sep. 8, 1999.

Newton MessagePad Handbook; Chapter 2, Handwriting, Text, and Drawing; 1993; pp. cover, i–viii, 1–39.

Automatic Speech Recognition; ASRNews; Market, Investment and Technical News of the Emerging Automatic Speech Recognition Industry; May 1992; vol. 3 No. 5; pp. cover, 6–7.

Automatic Speech Recognition; ASRNews; Market, Investment and Technical News of the Emerging Speech Recognition Industry; Sep 1992; vol. 3 No. 9; pp. cover, 6–7.

Richard A. Bolt; "Put-That-There": Voice and Gesture at the Graphics Interface; Architecture Machine Group Massachusetts Institute of Technology, Cambridge, Massachusetts; pp. 262–270.

Hewitt D. Crane and Dimitry Rtischev; State of the Art; Pen and Voice Unite; Adding pen and voice input to today's user interfaces opens the door for more natural communication with your computer; BYTE Oct. 1993; pp. title, 99.

Alex Rudnicky; State of the Art; Pen and Voice Unite; Matching the Input Mode of the Task; BYTE Oct. 1993; pp. 100–102.

Laurence Nigay and Joëlle Coutaz; Papers; "A Generic Platform for Addressing the Multimodal Challenge"; Abstract; May 7–11 1995 Chi '95 Mosaic of Creativity; pp. 98–105.

Sharon Oviatt, Antonella De Angeli and Karen Kuhn; "Integration and Synchronization of Input Modes During Multimodal Human—Computer Interaction"; Center for Human—Computer Communication, Department of Computer Science and Engineering, Oregon Graduate Institute of Science and Technology, Portland, OR; pp. 415–422.

Edwin Bos, Carla Huls and Wim Claassen, "EDWARD: full integration of language and action in a multimodal user interface"; Int. J. Human—Computer Studies (1994) 40, 1994 Academic Press Limited; pp. 473–495.

Will Hill, David Wroblewski, Tim McCandless, and Rich Cohen; "Architectural Qualities and Principles for Multimodal and Multimedia Interfaces"; Chapter 17; pp. 311–318.

Alexander I. Rudnicky and Alexander G. Hauptmann; "Multimodal Interaction in Speech Systems"; Chapter 10, pp. 147–171.

Amir Mane, Susan Boyce, Demetrios Karis, and Nicole Yankelovich; "Designing the User Interface for Speech Recognition Applications"; A Chi 96 Workshop; SIGCHI Bulletin; vol. 28, No. 4, Oct. 1996; pp. 29–34.

Alan Wexelblat; "Gesture at the User Interface"; SIGCHI Bulletin; Apr. 1996, vol. 28, No. 2, pp. 22–26.

Randy Pausch and Rich Gossweiler; "Application-Independent Object Selection From Inaccurate Multimodal Input"; Chapter 9; pp. 139–145.

Karl-Heinz Hanne and Hans-Jörg Bullinger; "Multimodal Communication: Integrating Text and Gestures"; Chapter 8; pp. 127–138.

M.L. Bourguet, S. Mimura, S. Ikeno and M. Komura; "A Multimodal Operational System for Security Services"; Intelligent Systems laboratory SECOM Co., Ltd; Symbiosis of Human and Artifact; 1995; pp. 219–224.

Philip R. Cohen, Michael Johnston, David McGee, Sharon Oviatt, Jay Pittman, Ira Smith, Liang Chen and Josh Clow; "QuickSet: Multimodal Interaction for Simulation Set-up and Control "Abstract; Center for Human Computer Communication, Oregon Graduate Institute of Science and Technology, Portland, OR; pp. 1–6.

Alex Waible, Minh Tue Vo, Paul Duchnowski, and Stefan Manke; "Multimodal Interfaces"; Abstract; Carnegie Mellon University, Pittsburgh, PA, 1996.

I. McKay, M.A. Jack, R. Thomson; "Report Generation using Multi-modal Data Entry for Office Systems"; Abstract; IEE Colloquium (Digest) Proceedings of the 1996 IEE Colloquium on Interfaces; 1996.

Stefan Muench, Ruediger Dillmann; "Haptic Output in Multimodal User Interfaces"; Abstract; International Conference on Intelligent User Interfaces, Proceedings of the 1997 International Conference on Intelligent User Interfaces; 1997.

Jean-Francois Arcand, Christopher Ramstein; "Artificial Neural Network for the Design of an Adaptive Multimodal Interface"; Abstract; Proceedings of the International Conference on Tools with Artificial Intelligence Proceedings of the 1995 IEEE 7th International Conference on Tools with Artificial Intelligence; 1995.

G. Faconti, M. Bordegoni, K. Kansy, P. Trahanias, T. Rist, and M. Wilson; "Formal Framework and Necessary Properties of the Fusion of Input Modes in User Interfaces"; Abstract; Interacting with Computers v 8 n 2 Jun. 1996, pp. 134–161.

Shiro Kawai, Hitoshi Aida and Tado Saito; "Designing Interface Toolkit with Dynamic Selectable Modality"; Abstract; Annual ACM Conference on Assistive Technologies, Proceedings of the 1996 2nd ACM Conference on assistive Technologies Apr. 11–12, 1996.

Alan Wexelblat and Marc Cavazza; "Gesture at the user Interface"; Abstract; Human Factors in Computing Systems (CHI)—Conference Proceedings Proceedings of the Conference on Human Factors in Computing Systems, May 7–11, 1995.

James Flanagan and Ivan Marsic; "Issues in Measuring the Benefits of Multimodal Interfaces"; Abstract; ICASSP, IEEE International Conference on acoustics, Speech and Signal Processing 1997.

Claudie Faure; "Pen and Voice Interface for Incremental Design of Graphic Documents"; Abstract IEE Colloquium (Digest) Computing and Control Division Colloquium on Handwriting and Pen-Based Input Mar. 11, 1994.

Y. Namba, S. Tano, and H. Kinukawa; "Semantic Analysis Using Fusionic Property of Multimodal Data"; Abstract; Transactions of the Information Processing Society of Japan, vol. 38, No. 7, pp. 1441–1453.

Shulin Yang and Kuo–Chu Chang; "Multimodal Pattern Recognition by Modular Neural Network"; Abstract; Optical Engineering, vol. 37, No. 2, pp. 650–659; Feb. 1998.

A. Burstein, A.C. Long, Jr., S. Narayanaswamy, R. Han and R.W. Brodersen; "The InfoPad User Interface"; Abstract; Digest of Papers COMPCON'95; 1995.

L. Nigay and J. Coutaz; "Design Spaces for Multimedia and Multimodal Interaction"; Abstract; Technique et Science Informatiques, vol. 15, No. 9, pp. 1195–1225; 1996.

S. Oviatt; "Multimodal Interactive Maps: Designing for Human Performance"; Abstract; Human—Computer Interaction, vol. 12, No. 1–2, pp. 93–129; 1997.

T. Wakahara, A. Suzuki, N. Nakajima, S. Miyahara and K. Odaka; "On–line Cursive Kanji Character Recognition as Stroke Correspondence Problem"; Abstract; Proceedings of the Third International Conference on Document Analysis and Recognition, p. 2, vol. xxvi +1188, 1059–64 vol. 2; 1995.

J.L. Leopold and A.L. Ambler; "Keyboardless Visual Programming Using Voice, Handwriting, and Gesture"; Abstract; Proceedings, 1997 IEEE Symposium on Visual Languages (Cat. No. 97TB100180), p. xiii+451, 28–35; 1997.

A. Malaviya, C. Leja and L. Peters; "Multi–script Handwriting Recognition with FOHDEL"; Abstract; 1996 Biennial Conference on the North American Fuzzy Information Processing Society—NAFIPS (Cat. No. 96th8171), p. 601, 147–51; 1996.

K. Kiyota, S. Yamamoto, N. Ezaki and T. Sakurai; "On–line Japanese Character Recognition System for Visually Disabled Persons"; Abstract, Proceedings of the 13th International Conference on pattern Recognition, p. 4; 1996.

Y. Shimada, M. Ohkura, M. Shiono and R. Hashimoto; "On Discrimination of Handwritten Similar Kanji Characters by Multiple Feature Subspace Method"; Abstract; Transactions of the Institute of Electronics, Information and Communications Engineers D–II, vol. J78D–II, No. 10, pp. 1460–1468; 1995 IEE; Oct. 1995.

K. Toyokawa, K. Kitamura, S. Katoh, H. Kaneko, N. Itoh and M. Fujita; "An On–line Character Recognition System for Effective Japanese Input"; Abstract; Proceedings of the Second International Conference on Document Analysis and Recognition (Cat. No. 93TH0578–5), pp. xx+963, 208–13; 1995.

Y. Bellik; "Media Integration in Multimodal Interfaces"; Abstract; 1997 IEEE First Workshop on Multimedia Signal Processing (Cat. No. 97TH8256), pp. xvi+596, 31–6; 1997.

M.T. Maybury; "Research on Multimedia and Multimodal Parsing and Generation"; Abstract; Artificial Intelligence Review, vol. 9, No. 2–3, pp. 103–127; Jun. 1995.

Y. Bellik; "The Time Component in Multimodal Interfaces"; Abstract; Genie Logiciel, No. 44, pp. 38–41; Jun. 1997.

A. Smith, J. Dunaway, P. Demasco and D. Peischi; "Multimodal Input for Computer Access and Augmentative Communication"; Abstract, ASSETS '96; The Second Annual ACM Conference on Assistive Technologies; p. 145, 80–5; 1996.

R. Sharma, T.S. Huang and V.I. Pavlovi'c; "A Multimodal Framework for Interacting with Virtual Environments"; Abstract; Human Interaction with Complex Systems: Conceptual Principles and Design Practice; p. 429, 53–71; 1996.

S. Oviatt and R. VanGent; "Error Resolution During Multimodal Human—computer Interaction"; Abstract; Proceedings ICSLP 96; Fourth International Conference on Spoken Language Processing (Cat. No. 96TH8206); p. 4 vol. 2522, 204–7 vol. 1; 1996.

W.M. Martinez; "A Natural Language Processor with Neural Networks"; Abstract; 1995 IEEE International Systems, Man and Cybernetics, Intelligent Systems for the 21st Centurey (Cat. No. 95CH3576–7), p. 5 vol. 4711, 3156–61 vol. 4; 1995.

Reality Fusion; Corporate Fact Sheet; 1997–1998.

Allan Christian Long, Jr., Shankar Narayanaswamy, Andrew Burstein, Richard Han, Ken Lutz, Brian Richards, Samuel Sheng, Robert W. Brodersen and Jan Rabaey; "A Prototype User Interface for a Mobile Multimedia Terminal"; Abstract; Department of Electrical Engineering and Computer Sciences; Chi '95 Proceedings.

"Computers and Software"; Popular Science; p. 31; Mar. 1998.

Shankar Narayanaswamy and Robert Brodersen; "Pen–and–Speech Based Circuit Schematic Editor".

http://www.cogsci.kun.nl/~miami/taxonomy/node1.html; Contents; printed Apr. 13, 1998.

Adam Cheyer and Luc Julia; "Multimodal Maps: An Agent–Based Approach"; Abstract; SRI International 1995.

Phillip Cohen et al.; "Synergistic Use of Direct Manipulation and Natural Language"; Abstract; SRI International 1989.

Sharon Oviatt and Erik Olsen; "Integration Themes in Multimodal Human—Computer Interaction"; Abstract; SRI International 1994.

Sharon Oviatt, Phillip Cohen, and Michelle Want; "Reducing Linguistic Variability in Speech and handwriting through Selection of Presentation Format"; Abstract; SRI International 1993.

Sharon Oviatt, Phillip Cohen, and Michelle Want; "Toward Interface Design for Human Language Technology: Modality and Structure as Determinants of Linguistic Complexity"; Abstract; SRI International 1994.

Catherine Wolf and Palmer Morrel-Samuels; "The Use of Hand-Drawn Gestures for Text Editing"; Abstract; Thomas J. Watson Research Center and Columbia University 1987.

Lynn Wilcox, Bill Schilit, and Nitin Sawhney; "Dynomite: A Dynamically Organized Ink and Audio Notebook"; Abstract; FX Palo Alto Laboratory and MIT 1997.

Jennifer Lai and John Vergo; "MedSpeak; Report Creatin with Continuous Speech Recognition"; Abstract, IBM and Thomas J. Watson Research Center 1997.

Scott Robertson, Cathleen Wharton, Catherine Ashworth; Marita Franzke; "Dual Device user Interface Design; PDAs and Interactive Television"; Abstract; Applied Research, Collaborative Systems Group, US West Advanced Technologies 1996.

Andrew Ortony, Jon Slack and Oliviero Stock; "Cognitive Science, Artificial Intelligence and Communication"; Abstract; Institute for the Learning Sciences, Northwestern University Instituto per la Ricerca Scientifica e Technologica 1995.

Wolfgang Wahlster, Elizabeth Andre, Som Bandyopadhyay, Winfried Graf, and Thomas Rist; "WIP: The Coordinated Generation of Multimodal Presentations from a Common Representation"; Abstract; German Research Center for Artificial Intelligence 1995.

Clive Frankish, Richard Hull, and Pam Morgan; "Recognition Accuracy and User Acceptance of Pen Interfaces"; Abstract; University of Bristol (U.K.), Hewlett Packard 1995.

Laurence Nigay, Joelle Coutaz; "A Generic Platform for Addressing the Multimodal Challenge"; Abstract; Laboratoire de Genie Informatique (LGI-IMAG) 1995.

T.V. Raman; "Emacspeak—A Speech Interface"; Abstract; Cambridge Research Lab, Digital Equipment Corp. 1996.

Bibliography of Papers; Spoken Language Interaction and Design; http://www.cse.ogi.edu/CHCC/Publications/...ltimodal Interaction and Interface Design; printed Nov. 18, 1997.

"Micro Power Hidden Markov Model State Decoders for Word-Spotting and Other Applications"; The University of California at Berkeley; Technology/Business Opportunity.

"Data-Voice Mobile Communication Sensor for Cooperative Automatic Vehicle Control"; The University of California at Berkely; Technology Business Opportunity.

"CAREER: Improving the Design of Interactive Software"; University of California; Office of Technology Transfer; Federally Funded Research Project.

Synthesis, Analysis, and Perception of Visible Speech; University of California; Office of Technology Transfer; Federally Funded Research Project.

Many People Consider the Futuristic; Hotlist Jan. 1998, pp. 7-11.

"Lets' Talk! Speech Technology is the Next Big Thing in Computing. Will it Put a PC in Every Home?"; Business Week; Feb. 23, 1998; pp. 61-80.

* cited by examiner

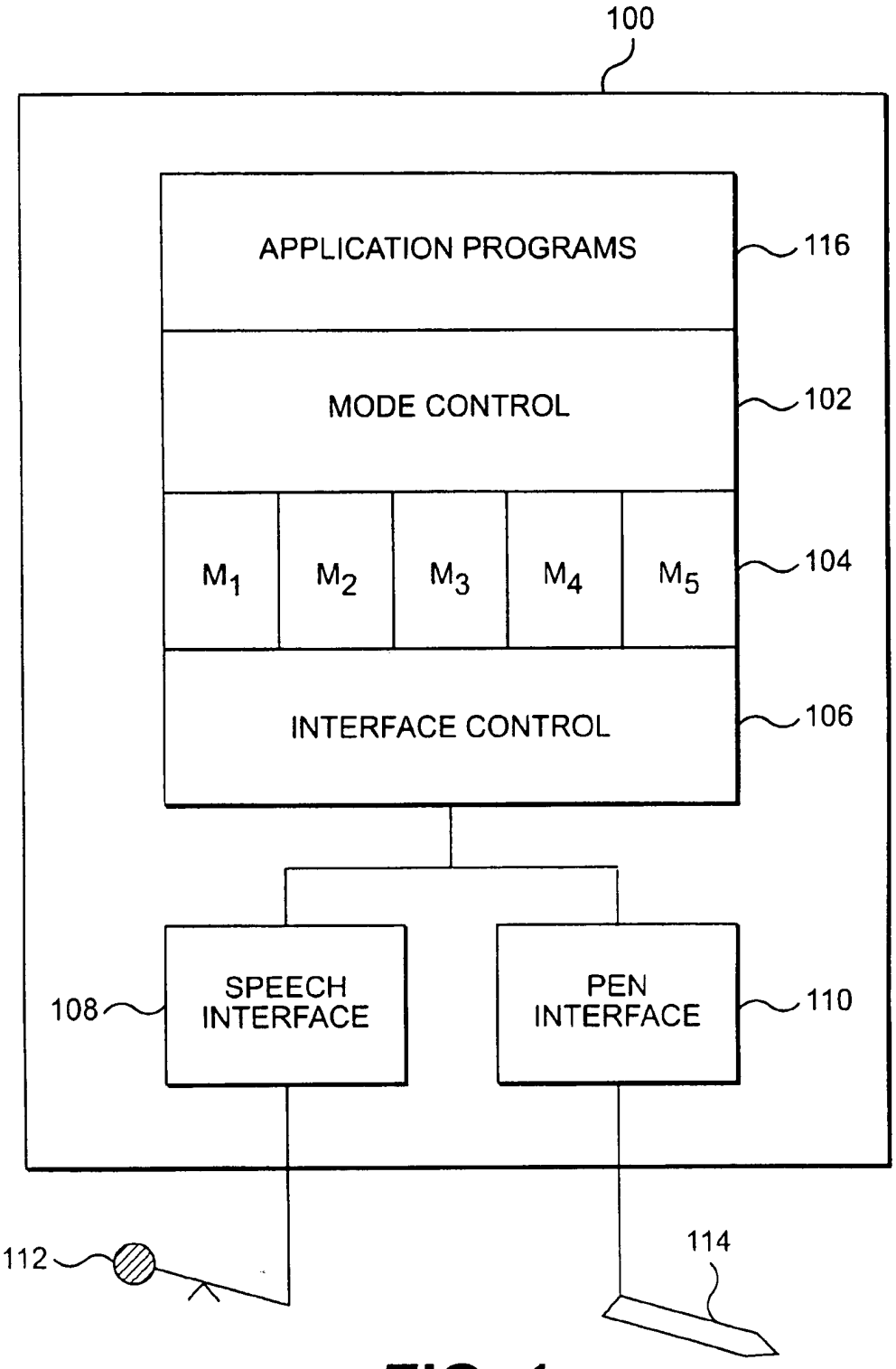

100
APPLICATION PROGRAMS
116
MODE CONTROL
102
M1
M2
M3
M4
M5
104
INTERFACE CONTROL
106
SPEECH INTERFACE
108
PEN INTERFACE
110
112
114

FIG. 1

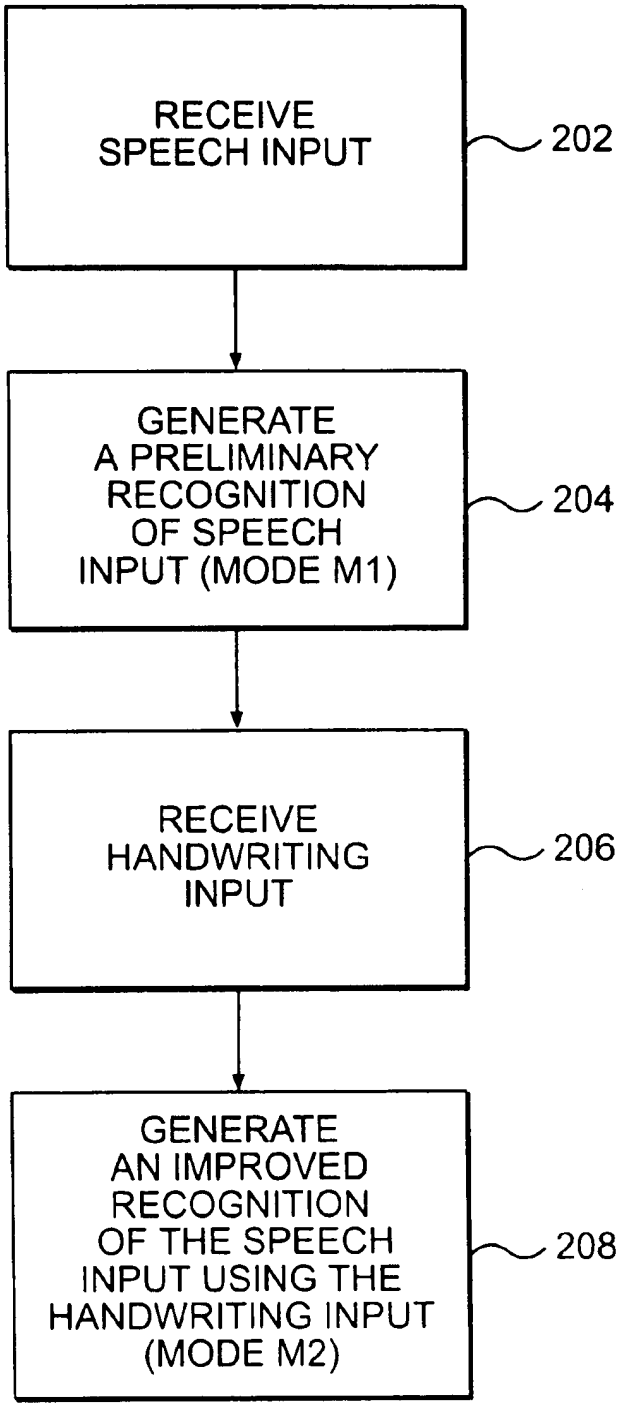

RECEIVE SPEECH INPUT
202
GENERATE A PRELIMINARY RECOGNITION OF SPEECH INPUT (MODE M1)
204
RECEIVE HANDWRITING INPUT
206
GENERATE AN IMPROVED RECOGNITION OF THE SPEECH INPUT USING THE HANDWRITING INPUT (MODE M2)
208

FIG. 2

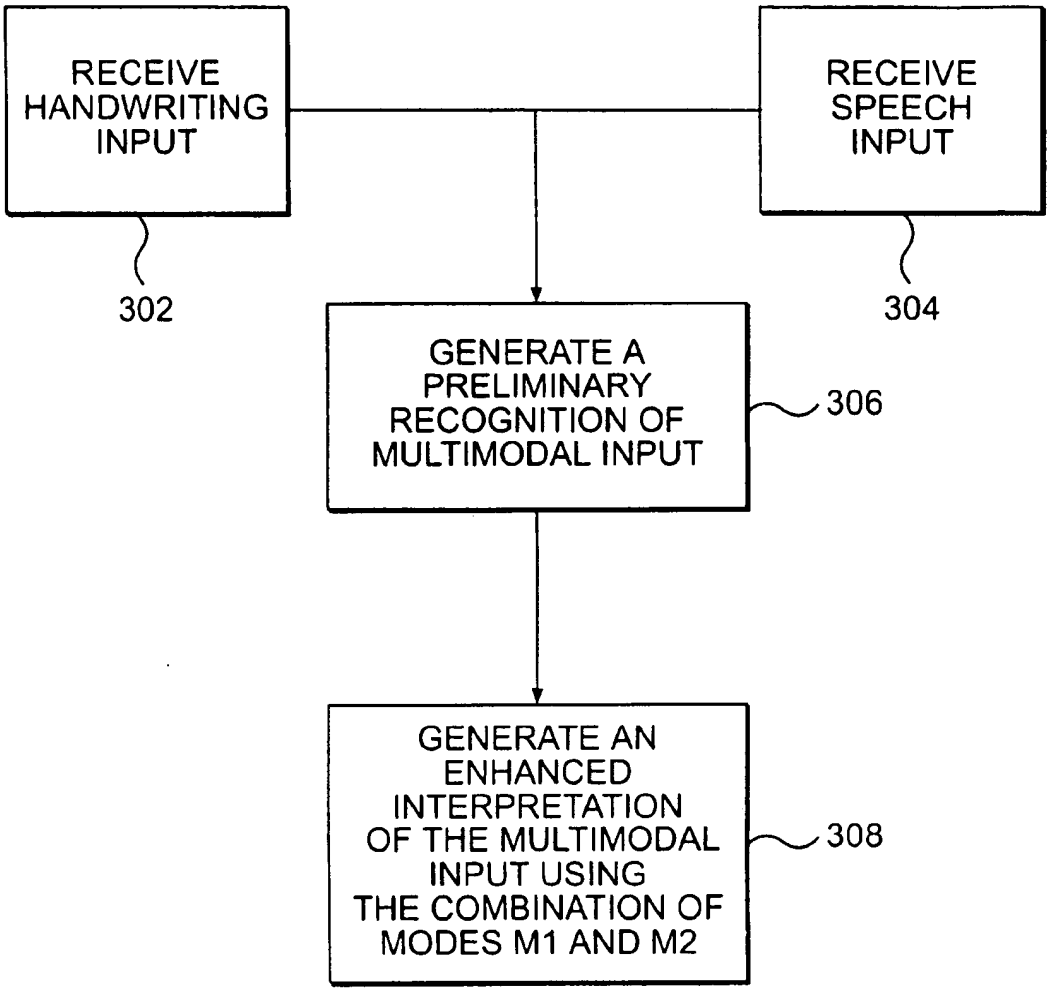

RECEIVE HANDWRITING INPUT
RECEIVE SPEECH INPUT
302
304
GENERATE A PRELIMINARY RECOGNITION OF MULTIMODAL INPUT
306
GENERATE AN ENHANCED INTERPRETATION OF THE MULTIMODAL INPUT USING THE COMBINATION OF MODES M1 AND M2
308

FIG. 3

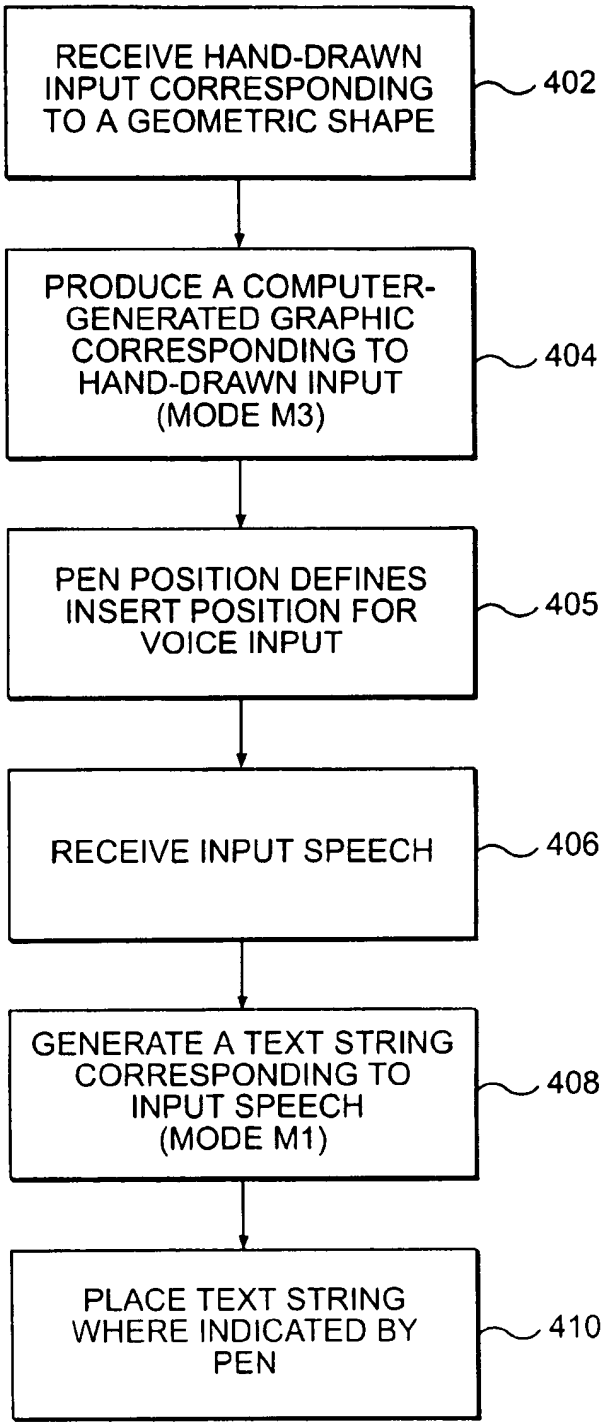

RECEIVE HAND-DRAWN INPUT CORRESPONDING TO A GEOMETRIC SHAPE
402
PRODUCE A COMPUTER-GENERATED GRAPHIC CORRESPONDING TO HAND-DRAWN INPUT (MODE M3)
404
PEN POSITION DEFINES INSERT POSITION FOR VOICE INPUT
405
RECEIVE INPUT SPEECH
406
GENERATE A TEXT STRING CORRESPONDING TO INPUT SPEECH (MODE M1)
408
PLACE TEXT STRING WHERE INDICATED BY PEN
410

FIG. 4

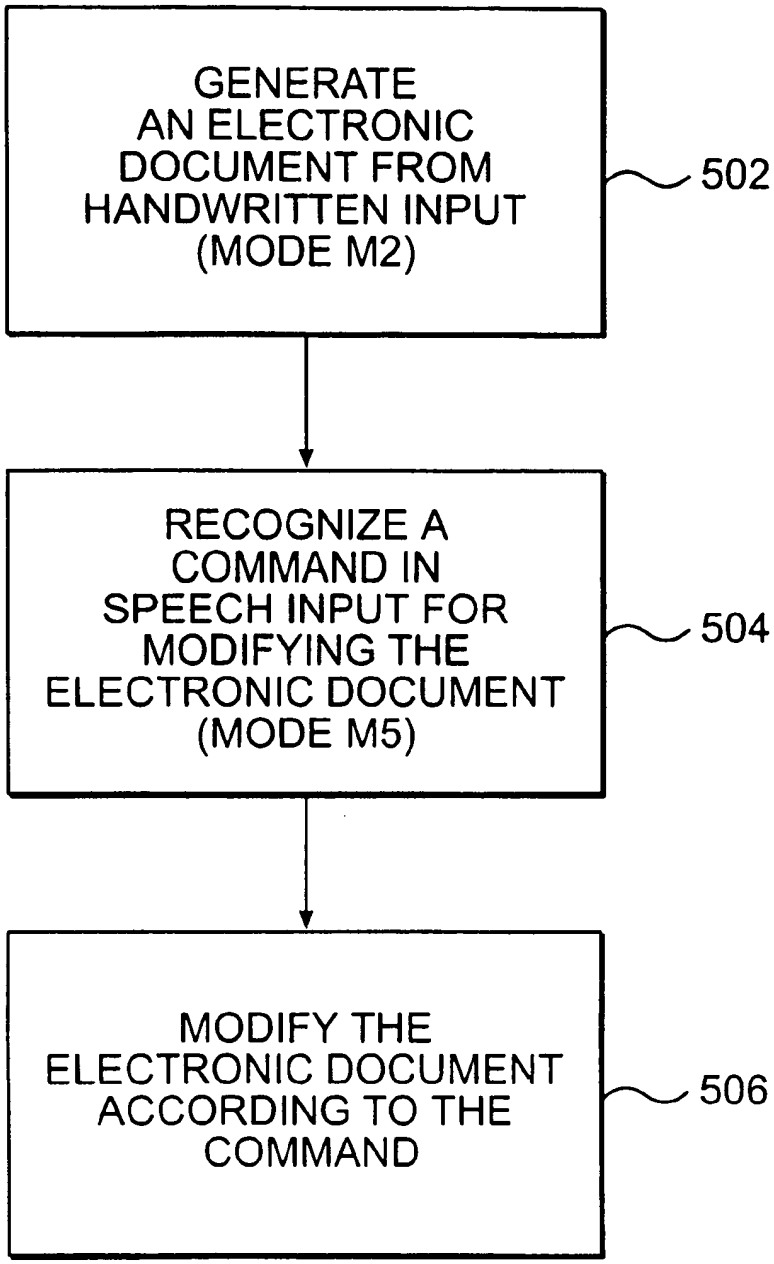

GENERATE AN ELECTRONIC DOCUMENT FROM HANDWRITTEN INPUT (MODE M2)
502
RECOGNIZE A COMMAND IN SPEECH INPUT FOR MODIFYING THE ELECTRONIC DOCUMENT (MODE M5)
504
MODIFY THE ELECTRONIC DOCUMENT ACCORDING TO THE COMMAND
506

FIG. 5

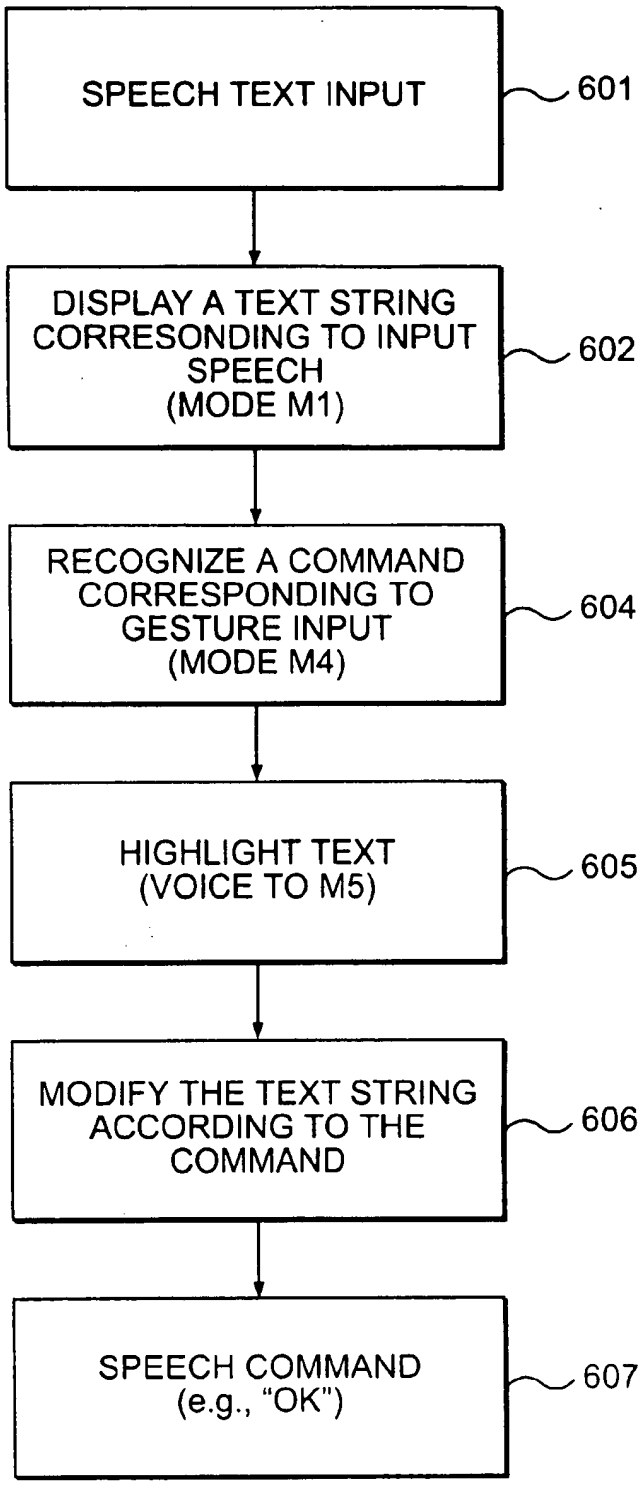

SPEECH TEXT INPUT
601
DISPLAY A TEXT STRING CORRESONDING TO INPUT SPEECH (MODE M1)
602
RECOGNIZE A COMMAND CORRESPONDING TO GESTURE INPUT (MODE M4)
604
HIGHLIGHT TEXT (VOICE TO M5)
605
MODIFY THE TEXT STRING ACCORDING TO THE COMMAND
606
SPEECH COMMAND (e.g., "OK")
607

FIG. 6

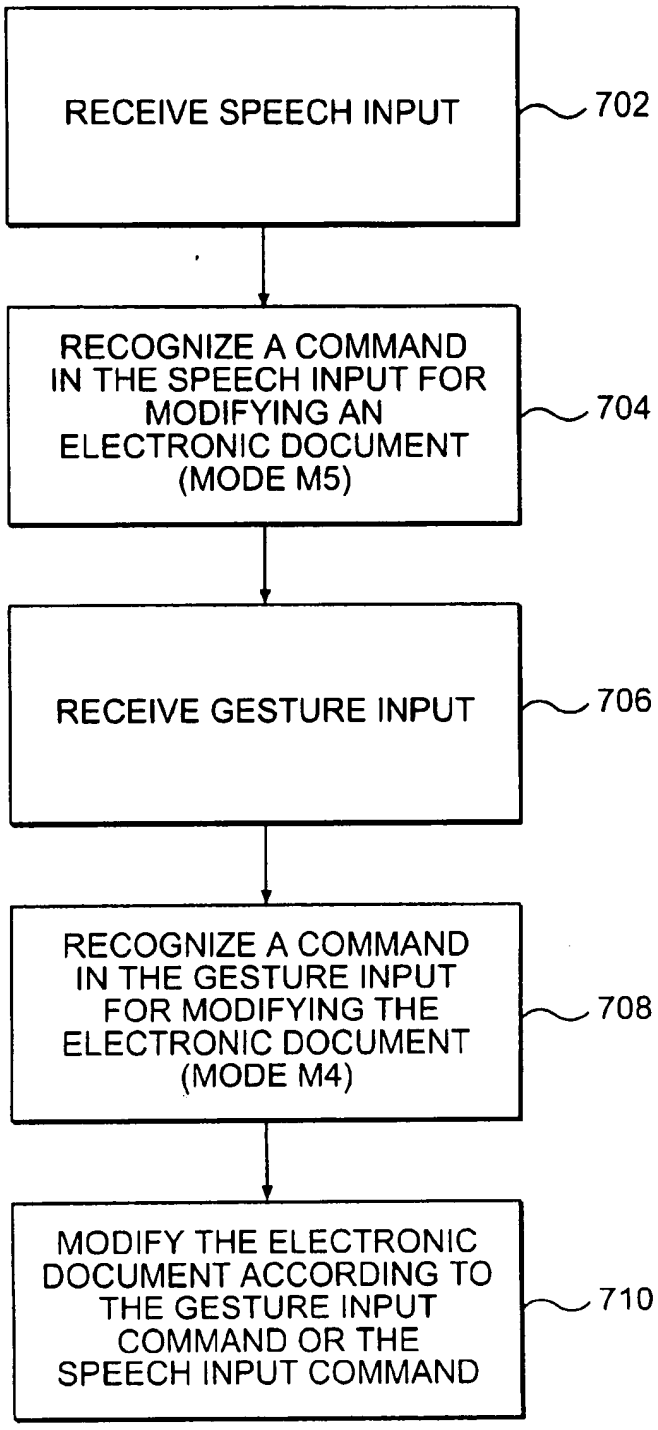

RECEIVE SPEECH INPUT
702
RECOGNIZE A COMMAND IN THE SPEECH INPUT FOR MODIFYING AN ELECTRONIC DOCUMENT (MODE M5)
704
RECEIVE GESTURE INPUT
706
RECOGNIZE A COMMAND IN THE GESTURE INPUT FOR MODIFYING THE ELECTRONIC DOCUMENT (MODE M4)
708
MODIFY THE ELECTRONIC DOCUMENT ACCORDING TO THE GESTURE INPUT COMMAND OR THE SPEECH INPUT COMMAND
710

FIG. 7

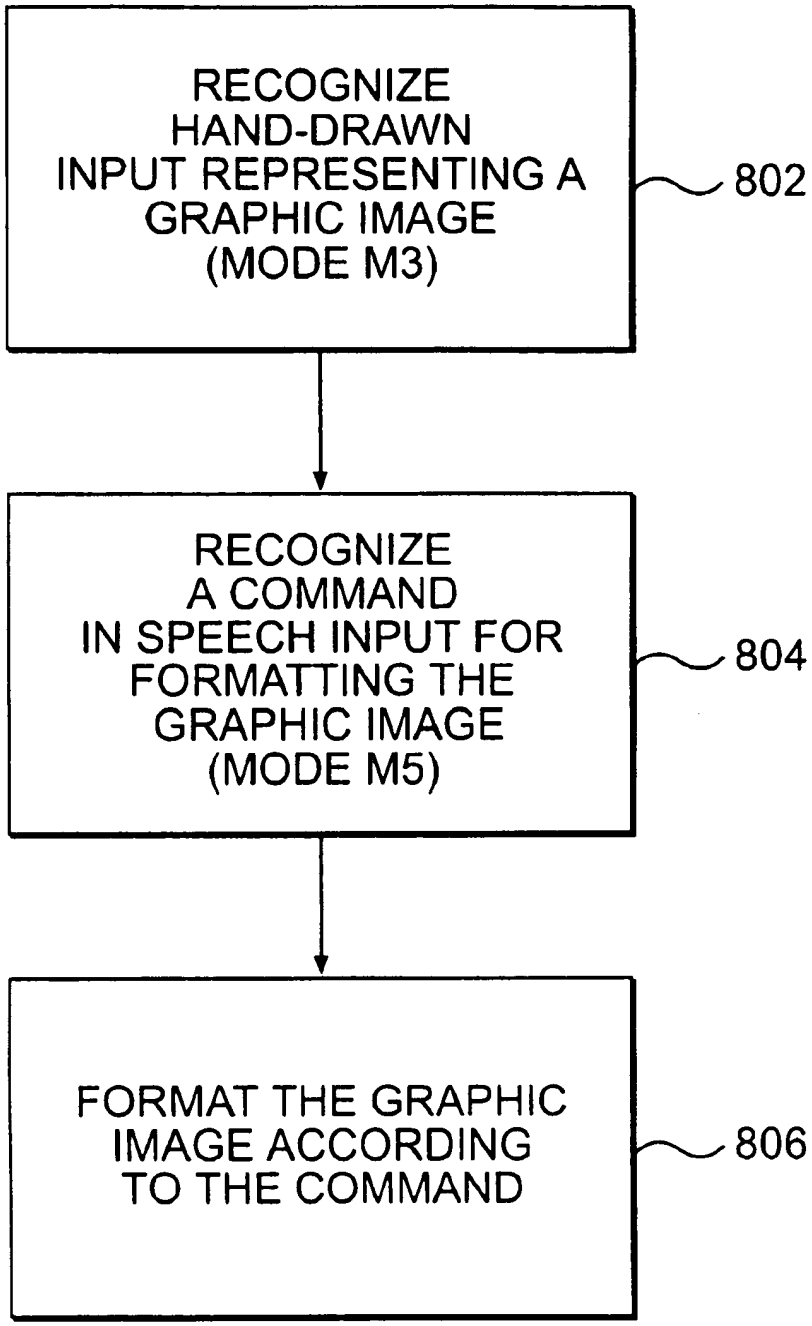

RECOGNIZE HAND-DRAWN INPUT REPRESENTING A GRAPHIC IMAGE (MODE M3)
802
RECOGNIZE A COMMAND IN SPEECH INPUT FOR FORMATTING THE GRAPHIC IMAGE (MODE M5)
804
FORMAT THE GRAPHIC IMAGE ACCORDING TO THE COMMAND
806

FIG. 8

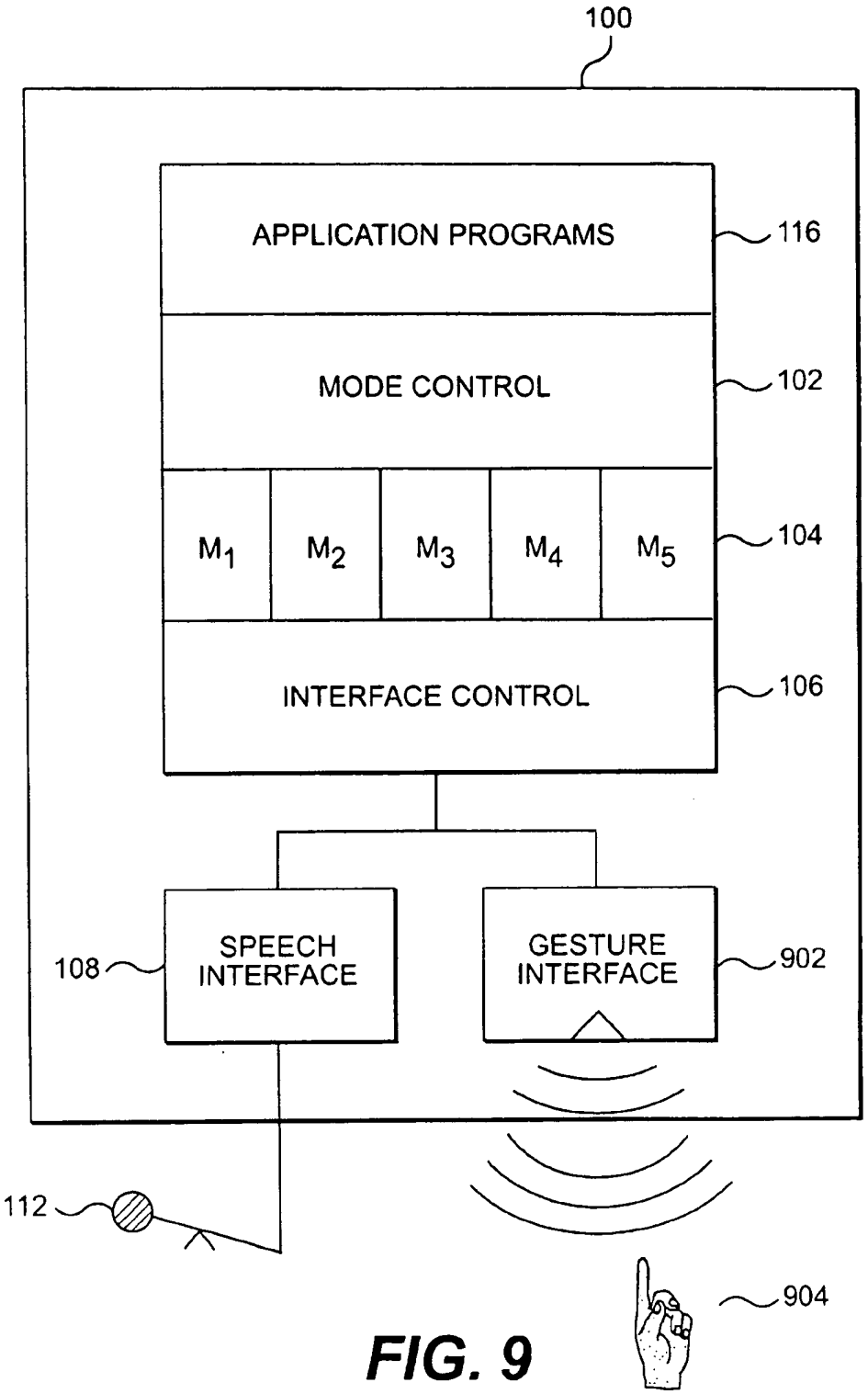

100
APPLICATION PROGRAMS
116
MODE CONTROL
102
M1
M2
M3
M4
M5
104
INTERFACE CONTROL
106
108
SPEECH INTERFACE
GESTURE INTERFACE
902
112
904

FIG. 9

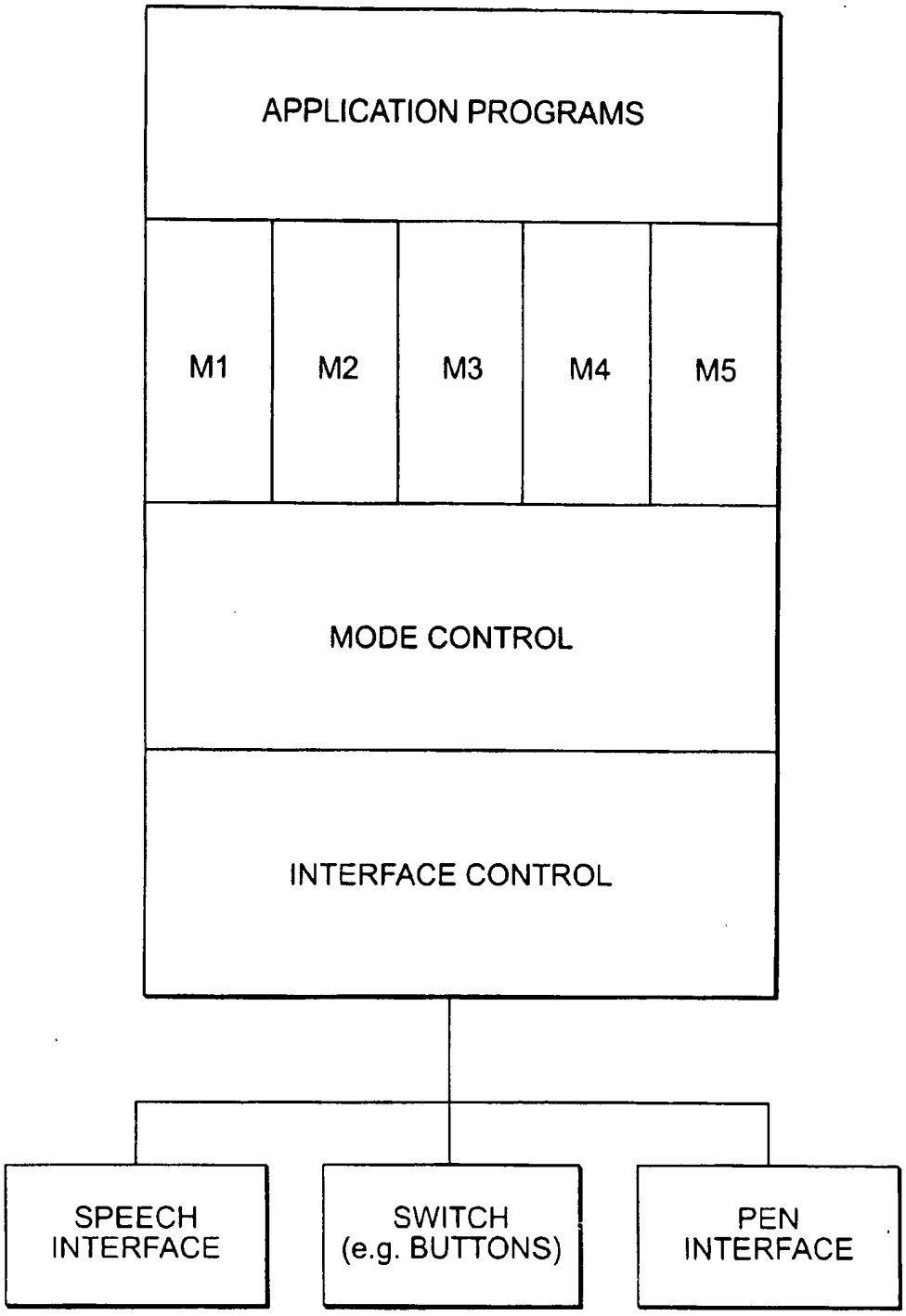

APPLICATION PROGRAMS
M1
M2
M3
M4
M5
MODE CONTROL
INTERFACE CONTROL
SPEECH INTERFACE
SWITCH (e.g. BUTTONS)
PEN INTERFACE

FIG. 10

# PROCESSING HANDWRITTEN AND HAND-DRAWN INPUT AND SPEECH INPUT

This application claims priority to U.S. Provisional Patent Application No. 60/086,346, filed May 20, 1998.

The ...al process-ing sy... to systems that pr... wn gesture input.

The ...ion entitled "APP... CESSING HANI... UT", Serial No. 6... icorporated by ref...

Since the early 1980s, personal computers have become increasingly powerful, able to store large amounts of information, create complex text, and multimedia documents including now color animation, 3D effects, and sound. In addition, these devices are able to communicate over telephone lines or local area networks with other computers directly or through the Internet. These computers are able to draw on large databases stored on large capacity hard drives in the personal computers. PCs can also tap into remote databases through their networking and communications capabilities.

Although the human interface to these computers has evolved to a certain extent from the early 1980s, in many ways the major element of this interface, the keyboard, is still very similar to that of a manual typewriter whose origins date to the late part of the 19th Century. For most computers, even in the mid-990s, the 100-key keyboard with alpha/numeric and function keys still forms the basic input means for accessing and creating information on personal and other computers. Ironically, the keyboard that is in common use has its basic layout designed to slow typists down. This design dates from the days of mechanical typewriters whose keys jammed when typists became too proficient. Although many people using computers have learned to type very rapidly, for many who do not learn to type well or who do not know how to type, the keyboard interface to the computer represents a barrier to its use. In addition, many people who do learn to type well can develop a repetitive stress disorder, an inflammation of the wrists which can result in the complete inability to type and therefore loss of productivity on the computer.

In the late 1980s a pointing device, called a mouse, was developed for computer input which allows the user to move a curser or indicator within the computer output display screen. By pointing and clicking a mouse, certain words or areas on the screen may be chosen by the user. In this way, navigation of the display screen and command of computer operations may be controlled by pointing to various items or words or icons on the screen. The pointing device may be a mouse, which indirectly points to items on the screen, or a pen-type device applied directly to the screen or even a finger with a special touch screen.

Other operations are possible using these devices such as highlighting a word in order to provide an additional command by means of other switches on the pointing device to delete the word or change its appearance. The development of the graphic user interfaces (GUI), have greatly enhanced the use of pointing devices for the human interface to the computer. Although these pointing devices may substitute for a series of keystrokes for moving a pointer around on the screen c...perations are basi... the key-board. I... ouse and keyboar... actical to use mo... ary commands t...

Since ...ech recognition ...come an increasi... er require significant computing power for their operation. Early speech recognition devices could be trained by an individual to respond to a small number of command words effectively substituting for command keys on the keyboard or a limited number of mouse clicks in a Windows interface. As computers have become more powerful in their computing speed and memory capacity, automatic speech recognition systems for computer input have become more capable. It is possible on personal computers to use voice input commands to activate any Windows command that appears in the menu structure using discrete or continuous speech recognition without requiring navigation through several layers of menus. Speech recognition systems are an especially powerful substitute for the keyboard for the input of individual words of text to create documents or for discrete commands. Such systems, however, are not a good substitute for the ease and speed of display screen navigation or other drawing operations (for example circling a block of text and moving it by dragging it to a new place on the screen), which can easily be provided by a mouse or other pointing device. Moreover, such speech recognition systems have difficulty determining whether the received speech is a command or text.

Although the promise of automatic speech recognition systems for text creation using computers is great because they are rapid and easy to use, these systems suffer from some significant limitations which have impeded their general use in computers. The accuracy of speech recognition systems, even those well trained to the voice of a single user, a... ...be significantly lo... ...ls outside of the vo... ...many business ap... ...ecognition systems te... ...us editing and fo... ...of punctuation m... ...mechanism for n... ...and carrying out th... ...g device which a... ...," highlighting v... ...ng and creating graphi... ...t.

The physical size of computers has limited their utility in certain applications. Like many electronic devices, computers have grown dramatically smaller as they have evolved. In recent years, laptop and even palmtop computers the size of small books have become popular. A computer the size of a book, which may be carried anywhere or a small pocket-sized device, has no room for a keyboard large enough to accommodate hands of most adults. In addition, if a computer is to be used in the field as a palmtop device or even in an airplane seat, the use of a mouse-type pointing device that requires an external pad is impractical. A pointing device such as a pen for use on even a small computer display surface is extremely useful.

A number of devices without keyboards have been proposed that use pens and have handwriting recognition as input and/or receive mouse-type input. Those introduced have had limited ability to recognize even fairly clear

handwriting. Although handwriting recognition by pen input devices has significantly improved in the last few years, like speech recognition, it still remains a challenging technical problem. For example, pen input in currently available systems is tiring and impractical when entering large amounts of text. Developing even smaller personal computing devices with the complete text input and computing capability of larger sized personal computers remains a major goal and interest of the computing public and the computing industry.

There is, therefore, a need for a computer system that departs from conventional methods and achieves increased performance by integrating speech recognition and handwritten and hand-drawn (e.g., pen or gesture input) recognition to overcome the disadvantages of either mode of recognition used alone or in an unintegrated combination.

## SUMMARY OF THE INVENTION

Methods and apparatus consistent with this invention process handwritten or hand-drawn input and speech input. Method steps include recognizing received handwritten or hand-drawn input, recognizing received speech input, and creating or modifying an electronic document according to the speech or handwritten or hand-drawn input.

An apparatus includes structure for recognizing handwritten or hand-drawn input, structure for recognizing speech input, and structure for activating modes for processing the handwritten or hand-drawn input and the speech input responsive to handwritten or hand-drawn input or the speech input.

Both the foregoing general description and the following detailed description are exemplary and explanatory and are intended to provide further explanation of the invention as claimed.

## DESCRIPTION OF THE FIGURES

The accompanying drawings provide a further understanding of the invention. They illustrate embodiments of the invention and, together with the description, explain the principles of the invention.

FIG. **1** is a block diagram of a computer system for processing handwritten or hand-drawn input and speech input.

FIG. **2** is a flow diagram of a method for processing speech and handwritten input;

FIG. **3** is a flow diagram of a second method for processing speech and handwritten input;

FIG. **4** is a flow diagram of a method for processing hand-drawn input and speech input;

FIG. **5** is a flow diagram of a method for processing handwritten input and speech input;

FIG. **6** is a flow diagram of a method for processing handwritten input and speech input;

FIG. **7** is a flow diagram of a method for editing an electronic document;

FIG. **8** is a flow diagram of a method for processing hand-drawn input and speech input;

FIG. **9** is a block diagram of a computer system for processing gesture, handwritten, or hand-drawn input and speech input; and

FIG. **10** is a block diagram of a computer system with a mode switch for processing gesture, handwritten, or hand-drawn input and speech input.

## DETAILED DESCRIPTION OF THE INVENTION

Reference will now be made in detail to preferred embodiments illustrated in the accompanying drawings. The same numbers in different figures refer to like or similar elements.

FIG. **1** is a block diagram of computer system **100** for processing handwritten or hand-drawn input and speech input comprising mode controller **102**, mode processing logic **104**, interface controller **106**, speech interface **108**, pen interface **110**, and application programs **116**. Microphone **112** connects to speech interface **108**. Electronic pen **114** connects to pen interface **110**.

Interface controller **106** controls speech interface **108** and pen interface **110**, providing pen or speech input to mode controller **102**. Speech interface **108** preferably includes computer hardware and software for encoding an electronic signal generated by microphone **112** into a digital stream for processing by mode processing logic **104**. Similarly, pen interface **110** preferably includes computer hardware and software for processing handwritten or hand-drawn input created with electronic pen **114**.

Mode controller **102** activates modes in mode processing logic **104** according to input received from interface controller **106** to create an operating state for computer system **100**. An operating state governs how input received from interface controller **106** is processed and passed to application programs **116**. Application programs **116** include, for example computer programs for creating, editing, and viewing electronic documents, such as word processing, graphic design, spreadsheet, electronic mail, and web browsing programs.

An operating state is defined by one or more active modes in mode processing logic **104**. Mode processing logic **104** preferably supports five modes. The processing logic in mode **M1** recognizes speech input to microphone **112**. In mode M1, computer system **100** translates the speech input into a machine encoded (e.g., ASCII) text stream. For example, in mode M1, computer system **100** translates continuous speech spoken by an operator into microphone **112** into machine encoded text data forming part of an electronic document.

In mode M2, the processing logic recognizes pen input, i.e., it translates handwritten or hand-drawn input received from electronic pen **114** into a machine encoded text stream. For example, In mode M2, computer system **100** translates hand-printed or cursive writing written by an operator using electronic pen **114** into machine encoded text data forming part of an electronic document.

In mode M3, the processing logic recognizes hand-drawn graphic images created with electronic pen **114**. For example, in mode M3, computer system **100** recognizes an attempt by the operator to draw a rough outline for a circle, triangle, or square. The processing logic in mode may M3 create a corresponding "clean" (e.g., smooths curves, straightens lines, and corrects any geometric distortions) electronic representation of the drawing by issuing a snap-to-grid command.

In mode M4, the processing logic treats input received from electronic pen **114** as a command or an indication of a position in an electronic display. In mode M5, the processing logic recognizes speech input received from microphone **112** as a command: The commands recognized in modes M4 and M5 include, for example, commands or menus that are accessible from tool bars and pull-down menus in many graphics and text application programs.

A "combination mode" consists of two or more active modes. One such combination mode is an operating state where modes M1 and M2 are active. FIG. **2** is a flow diagram of a method for processing speech and handwritten input. According to the method shown in FIG. **2**, the computer system receives speech input by an operator (step

202). Next the computer system generates a preliminary recognition of the speech input using mode M1 (step 204). For example, the computer's preliminary classification would identify several text strings, letters, characters, or words corresponding to the closest matches to the recognized speech input. The operator then enters handwriting input providing the computer system with additional data to improve recognition of the input speech (step 206). Using mode M2, computer system 100 generates an improved recognition of the input speech from additional data in mode M2 to select from among recognition candidates identified by mode M1 at step 204.

FIG. 3 is a flow diagram of a method for processing speech and handwritten input where speech input is used to enhance the interpretation of handwriting or hand-drawn input and speech input (multimodal input). According to the method shown in FIG. 3, computer system 100 receives handwriting input (step 302) and speech input from an operator (step 304). Next system 100 generates a preliminary recognition of the multimodal input (step 306). Using mode M1 and M2, the computer system generates an enhanced interpretation of the input using the additional data provided by both the handwritten or hand-drawn input and input speech (step 308).

In a second combination mode, the operating state has both modes M1 and M3 active. FIG. 4 is a flow diagram of a method for processing hand-drawn input and speech input. Computer system 100 receives a geometric shape or figure drawn by the operator using electronic pen 114 (step 402). System 100 generates a graphic corresponding to the hand-drawn geometric shape using processing mode M3 (step 404). The location of the electronic pen defines an insert position in an electronic document for text recognized from the speech input (step 405). The computer system then receives speech input corresponding to text to be placed in the electronic document (step 406). Using processing mode M1, computer system 100 generates a text string from the speech input (step 408). System 100 places the text where indicated by the pen (step 410). The foregoing method is useful when creating annotated charts, for example, corporate organization charts. Using this method, an operator uses an electronic pen to sketch rough outlines of boxes corresponding to the organization structure, the computer cleans-up the boxes, and the user dictates annotations that are automatically inserted where indicated by the pen in the figure.

In a third combination mode, the operating state has both modes M2 and M5 active. FIG. 5 is a flow diagram of a method for processing handwritten input and speech input. Using a computer input device, for example, electronic pen 114, an operator writes text that computer system 100, using processing logic in mode M2, recognizes and generates an electronic document (step 502). To format or modify the document, the operator speaks commands into microphone 112 that are recognized by computer system 100 using processing logic in mode M5 (step 504). Computer system 100 modifies the electronic document as instructed by the command (step 506).

Using computer system 100 in an operating state with modes M2 and M5 active, an operator can create a memo by printing or cursive writing and have the words recognized and displayed on a computer screen. To format the memo as it is being created, the operator can utter commands into the microphone such as "bold," "underline," "paragraph," etc. The electronic document that is displayed on the screen will change in response to these commands.

In a fourth combination mode, the operating state has both modes M1 and M4 active. FIG. 6 is a flow diagram of a method for processing handwritten input and speech input. In this operating state, mode M1 processing logic recognizes speech spoken into microphone 114 (step 601) and displays a corresponding text string on a computer display (not shown) connected to computer system 100 (step 602). The operator using electronic pen 114 makes gestures, for example, directly on the display or on an electronic tablet that are recognized by M4 processing logic as commands for modifying the text (step 604). (Any voice commands would be handled by M5 processing logic when the M5 mode is activated.) Computer system 100 then applies the command to the text in accordance with an established mapping of gestures to commands (step 606). Examples of these gesture-to-command mappings include circling a word to indicate highlighting, which activates combination mode M4–M5, thereby switching the system from state M1–M5 to state M4–M5. A spoken command such as "bold" accomplishes the editing. A spoken command, such as "OK" completes the process (step 607). Another example of gesture-to-command mapping is scratching through a word on the screen with the pen to delete it, activating state M1–M2 that allows a new word to be written in place of the scratched out word.

In a fifth combination mode, the operating state has both modes M4 and M5 active. FIG. 7 is a flow diagram of a method for editing an electronic document in this operating state. Computer system 100 receives speech input from an operator specifying an editing command (step 702). Using M5 processing logic, computer system 100 recognizes a command in the speech input (step 704). Computer system 100 also receives gesture input from the operator indicating a command that can also include the spatial coordinates of the portion of the electronic document to be edited (step 706). Computer system 100 M4 processing logic recognizes a command in the gesture input (step 708). Computer system 100 then modifies the electronic document according to either or both the command from the speech input and the command from the gesture input (step 710).

Using computer system 100 in this operating state is advantageous when an operator desires to modify an electronic document, such as a drawing created in a computer graphics application program or a document created in a word processing application program. For example, to modify a drawing in a graphics editing context using electronic pen 114, the operator would circle a portion of the drawing displayed on a computer screen. The operator would then say "change color from red to blue." Application program 116 then responds to the spoken and gesture commands and changes the color of the circled region in the drawing from red to blue.

In a sixth combination mode, the operating state has both modes M3 and M5 active. FIG. 8 is a flow diagram of a method for processing hand-drawn input and speech input. Using processing logic in mode M3, Computer system 100 recognizes hand-drawn input from electronic pen 114 corresponding to, for example, graphic images (circles, triangles, squares, line drawings, etc.) (step 802). Mode processing logic in mode M5 recognizes a command in the speech input for formatting the graphic image (step 804). Application program 116 formats the graphic image according to the command (step 806). This operating state allows an operator to create drawings quickly without navigating through complex pull-down menus. For example, an operator sketches the outline of a square with electronic pen 114 and says "fill color black, cross-hatch, dashed outline." Computer system 100 then creates a square (straightens the lines drawn by the operator, makes them of equal length, and

connects them at right angles) and formats it according to the spoken commands.

The foregoing five modes are not exhaustive of all possible modes. Moreover, the six combinations of modes is only illustrative of a limited number of the many possible combinations of the modes for an operating state. Therefore, this invention is not limited to the particular modes and combinations disclosed, but includes all modes and combinations falling within the scope of the claims.

In an embodiment, computer system **100** mode controller **102** switches between modes and combination modes thereby switching operating states. Switching operating states changes the manner in which structure for recognizing gesture input recognizes gestures, handwritten, and hand-drawn input, and the structure for recognizing speech input recognizes speech. For example, as described in greater detail above, switching among modes and combinations of modes will control whether the recognition process will treat gesture input as a command, text, or drawing, and whether speech input is recognized as a command or text. Switching among operating states can be controlled by, for example, application program **116**, gestures received by pen interface **110**, speech received by speech interface **108**, or an electronic or mechanical switch. A switch or button can be connected to computer system **100**, microphone **112**, electronic pen **114**, or any other peripheral device associated with computer system **100** (e.g., a digitizing tablet connected to computer system to control switching among operating states. Other types of controls for switching operating states include rotating wheels, numeric keypads, and chorded keypads for one-handed letter input. FIG. **10** is a block diagram of a computer system with switch **1002** for processing handwritten and hand-drawn input and speech input.

Methods and apparatus for processing speech and handwritten and hand-drawn input are suitable for several application environments including, but not limited to, information kiosks, television/video cassette recorder remote control, a low profile computer with a form factor similar to a pad of paper (e.g., slate or pen tablet computer), a palm computer, a telephone, an electronic whiteboard, or a handheld personal computer.

Computer system **100** can receive several types of devices for providing gesture interaction with mode processing logic **104**. FIG. **9** is a block diagram of computer system **100** for processing gesture, handwritten, and hand-drawn input and speech input comprising several of the elements described in greater detail above in FIG. **1**, and wireless gesture interface **902**. Pointing device **904** is an electronic glove that tracks and transmits movement of the operator's hand. Alternatively, pointing device **904** could be a three-dimensional wireless mouse or wand. In yet another implementation, structure for providing gesture interface **902** includes a video tracking device, for example, the interface made by Reality Fusion that tracks the movement of the operator's hand as it makes gestures in the air.

A computer display can be connected to, or integrated within, computer system **100**. To provide the operator with feedback regarding the operating state, the appearance of the display changes in accordance with a change in operating state. For example, in a combination mode with M1 and M4 active, the display would have a white background. Changing the operating state to a combination mode with M2 and M5 active would display ruled or grid lines on the background of the display. Similarly, a background color change would indicate yet another operating state.

While some embodiments have been described, various changes and modifications may be made, and equivalents may be substituted. In addition, many modifications may be made to adapt a particular element, technique or implementation to the teachings of the present invention without departing from the central scope of the invention. For example, disclosed elements may be implemented in hardware, computer program code, or a combination of both hardware and computer program code. Moreover, elements depicted and described separately may be combined and implemented in a single element or distributed across a computer network. Therefore, this invention is not limited to the particular embodiments and methods disclosed, but includes all embodiments falling within the scope of the appended claims.

We claim:

**1**. A method for processing handwritten input or hand-drawn input and speech input comprising the steps, performed by a computer, of:

selecting a first operating mode and a second operating mode before processing handwritten input, hand-drawn input, or speech input;

recognizing the speech input as either a command or text for an electronic document in accordance with said first operating mode;

recognizing the handwritten or hand-drawn input as a command, text for an electronic document, or a graphic for an electronic document in accordance with said second operating mode; and

switching said first operating mode or said second operating mode to change how the computer recognizes said speech input or said handwritten or hand-drawn input.

**2**. The method of claim 1, wherein the step of switching said first operating mode or said second operating mode includes the substep of switching either said first mode or said second mode in response to a gesture.

**3**. The method of claim 1, wherein the step of switching said first operating mode or said second operating mode includes the substep of switching either said first mode or said second mode in response to speech input.

**4**. The method of claim 1, wherein the step of switching said first operating mode or said second operating mode includes the substep of switching either said first mode or said second mode in response to a mechanical switch activated by an operator.

**5**. A method for processing handwritten input or hand-drawn input and speech input comprising the steps, performed by a computer, of:

selecting a first combination mode and a second combination mode before processing handwritten hand-drawn, or speech input;

processing said handwritten input or said speech input in said first combination mode wherein said speech input is recognized as text input and said handwritten input is recognized as a command;

processing said handwritten input or said speech input in said second combination mode wherein said speech input is recognized as text and said handwritten input is recognized as text; and

switching processing modes between said first combination mode and said second combination mode.

**6**. A method for processing handwritten input or hand-drawn input and speech input comprising the steps, performed by a computer, of:

selecting a first combination mode and a second combination mode before processing handwritten, hand-drawn, or speech input;

processing said handwritten input or said speech input in said first combination mode wherein said speech input is recognized as text input and said handwritten input is recognized as a command;

processing said handwritten input or said speech input in said second combination mode wherein said speech input is recognized as a command and said handwritten input is recognized as a command; and

switching processing modes between said first combination mode and said second combination mode.

7. A method for processing handwritten input or handdrawn input and speech input comprising the steps, performed by a computer, of:

selecting a first combination mode and a second combination mode before processing handwritten, handdrawn, or speech input;

processing said handwritten input or said speech input in said first combination mode wherein said speech input is recognized as text input and said handwritten input is recognized as a command;

processing said handwritten input or said speech input in said second combination mode wherein said speech input is recognized as a command and said handwritten input is recognized as text; and

switching processing modes between said first combination mode and said second combination mode.

8. A method for processing handwritten input or handdrawn input and speech input comprising the steps, performed by a computer, of:

selecting a first combination mode and a second combination mode before processing handwritten, handdrawn, or speech input;

processing said handwritten input or said speech input in said first combination mode wherein said speech input is recognized as text and said handwritten input is recognized as text;

processing said handwritten input or said speech input in said second combination mode wherein said speech input is recognized as a command and said handwritten input is recognized as a command; and

switching processing modes between said first combination mode and said second combination mode.

9. A method for processing handwritten input or handdrawn input and speech input comprising the steps, performed by a computer, of:

selecting a first combination mode and a second combination mode before processing handwritten, handdrawn, or speech input;

processing said handwritten input or said speech input in said first combination mode wherein said speech input is recognized as text and said handwritten input is recognized as text;

processing said handwritten input or said speech input in said second combination mode wherein said speech input is recognized as a command and said handwritten input is recognized as text; and

switching processing modes between said first combination mode and said second combination mode.

10. A method for processing handwritten input or handdrawn input and speech input comprising the steps, performed by a computer, of:

selecting a first combination mode and a second combination mode before processing handwritten, handdrawn, or speech input;

processing said handwritten input or said speech input in said first combination mode wherein said speech input is recognized as a command and said handwritten input is recognized as text;

processing said handwritten input or said speech input in said second combination mode wherein said speech input is recognized as a command and said handwritten input is recognized as a command; and

switching processing modes between said first combination mode and said second combination mode.

11. A method for processing handwritten input or handdrawn input and speech input comprising the steps, performed by a computer, of:

selecting a first combination mode and a second combination mode before processing handwritten, handdrawn, or speech input;

processing said handwritten input or said speech input in said first combination mode wherein said speech input is recognized as text input and said handwritten input is recognized as a command;

processing said handwritten input or said speech input in said second combination mode wherein said speech input is recognized as text input and said handwritten input is recognized as a graphic; and

switching processing modes between said first combination mode and said second combination mode.

12. A method for processing handwritten input or handdrawn input and speech input comprising the steps, performed by a computer, of:

selecting a first combination mode and a second combination mode before processing handwritten, handdrawn, or speech input;

processing said handwritten input or said speech input in said first combination mode wherein said speech input is recognized as text input and said handwritten input is recognized as a command;

processing said handwritten input or said speech input in said second combination mode wherein said speech input is recognized as a command and said handwritten input is recognized as a graphic; and

switching processing modes between said first combination mode and said second combination mode.

13. A method for processing handwritten input or handdrawn input and speech input comprising the steps, performed by a computer, of:

selecting a first combination mode and a second combination mode before processing handwritten, handdrawn, or speech input;

processing said handwritten input or said speech input in said first combination mode wherein said speech input is recognized as text and said handwritten input is recognized as text;

processing said handwritten input or said speech input in said second combination mode wherein said speech input is recognized as text input and said handwritten input is recognized as a graphic; and

switching processing modes between said first combination mode and said second combination mode.

14. A method for processing handwritten input or handdrawn input and speech input comprising the steps, performed by a computer, of:

selecting a first combination mode and a second combination mode before processing handwritten, handdrawn, or speech input;

processing said handwritten input or said speech input in said first combination mode wherein said speech input is recognized as text and said handwritten input is recognized as text;

processing said handwritten input or said speech input in said second combination mode wherein said speech input is recognized as a command and said handwritten input is recognized as a graphic; and

switching processing modes between said first combination mode and said second combination mode.

15. A method for processing handwritten input or hand-drawn input and speech input comprising the steps, performed by a computer, of:

selecting a first combination mode and a second combination mode before processing handwritten, hand-drawn, or speech input;

processing said handwritten input or said speech input in said first combination mode wherein said speech input is recognized as text input and said handwritten input is recognized as a graphic;

processing said handwritten input or said speech input in said second combination mode wherein said speech input is recognized as a command and said handwritten input is recognized as a command; and

switching processing modes between said first combination mode and said second combination mode.

16. A method for processing handwritten input or hand-drawn input and speech input comprising the steps, performed by a computer, of:

selecting a first combination mode and a second combination mode before processing handwritten, hand-drawn, or speech input;

processing said handwritten input or said speech input in said first combination mode wherein said speech input is recognized as text input and said handwritten input is recognized as a graphic;

processing said handwritten input or said speech input in said second combination mode wherein said speech input is recognized as a command and said handwritten input is recognized as text; and

switching processing modes between said first combination mode and said second combination mode.

17. A method for processing handwritten input or hand-drawn input and speech input comprising the steps, performed by a computer, of:

selecting a first combination mode and a second combination mode before processing handwritten, hand-drawn, or speech input;

processing said handwritten input or said speech input in said first combination mode wherein said speech input is recognized as text input and said handwritten input is recognized as a graphic;

processing said handwritten input or said speech input in said second combination mode wherein said speech input is recognized as a graphic; and

switching processing modes between said first combination mode and said second combination mode.

18. A method for processing handwritten input or hand-drawn input and speech input comprising the steps, performed by a computer, of:

selecting a first combination mode and a second combination mode before processing handwritten, hand-drawn, or speech input;

processing said handwritten input or said speech input in said first combination mode wherein said speech input is recognized as a command and said handwritten input is recognized as a command;

processing said handwritten input or said speech input in said second combination mode wherein said speech input is recognized as a command and said handwritten input is recognized as a graphic; and

switching processing modes between said first combination mode and said second combination mode.

19. A method for processing handwritten input or hand-drawn input and speech input comprising the steps, performed by a computer, of:

selecting a first combination mode and a second combination mode before processing handwritten, hand-drawn, or speech input;

processing said handwritten input or said speech input in said first combination mode wherein said speech input is recognized as a command and said handwritten input is recognized as text input;

processing said handwritten input or said speech input in said second combination mode wherein said speech input is recognized as a command and said handwritten input is recognized as a graphic; and

switching processing modes between said first combination mode and said second combination mode.

20. A pen tablet computer for processing handwritten input or hand-drawn input and speech input comprising:

selecting a first operating mode and a second operating mode before processing handwritten, hand-drawn, or speech input;

means for recognizing the speech input as either a command or text for an electronic document in accordance with said first operating mode;

means for recognizing the handwritten or hand-drawn input as a command, text for an electronic document, or a graphic for an electronic document in accordance with said second operating mode; and

means for switching said first operating mode or said second operating mode to change how the computer recognizes said speech input or said handwritten or hand-drawn input.

21. The pen tablet computer of claim 20 further comprising a pen-integrated microphone.

22. A palm computer for processing handwritten input or hand-drawn input and speech input comprising:

selecting a first operating mode and a second operating mode before processing handwritten, hand-drawn, or speech input;

means for recognizing the speech input as either a command or text for an electronic document in accordance with said first operating mode;

means for recognizing the handwritten or hand-drawn input as a command, text for an electronic document, or a graphic for an electronic document in accordance with said second operating mode; and

means for switching said first operating mode or said second operating mode to change how the computer recognizes said speech input or said handwritten or hand-drawn input.

23. The palm computer of claim 22 further comprising a pen-integrated microphone.

24. A kiosk computer system for processing handwritten input or hand-drawn input and speech input comprising:

selecting a first operating mode and a second operating mode before processing handwritten, hand-drawn, or speech input;

means for recognizing the speech input as either a command or text for an electronic document in accordance with said first operating mode;

means for recognizing the handwritten or hand-drawn input as a command, text for an electronic document, or a graphic for an electronic document in accordance with said second operating mode; and

means for switching said first operating mode or said second operating mode to change how the computer recognizes said speech input or said handwritten or hand-drawn input.

25. The kiosk computer system of claim 24, wherein said means for recognizing the handwritten or hand-drawn input includes either a touch screen, a digitizing tablet, or a video gesture recognizer.

26. A computer for processing handwritten input or hand-drawn input and speech input comprising:

selecting a first operating mode and a second operating mode before processing handwritten, hand-drawn, or speech input;

means for recognizing the speech input as either a command or text for an electronic document in accordance with said first operating mode;

means for recognizing the handwritten or hand-drawn input as a command, text for an electronic document, or a graphic for an electronic document in accordance with said second operating mode; and

means for switching said first operating mode or said second operating mode to change how the computer recognizes said speech input or said handwritten or hand-drawn input.

27. The computer of claim 26, wherein said means for recognizing the handwritten or hand-drawn input includes either a touch screen, a digitizing tablet, a three-dimensional wireless mouse, or a video gesture recognizer.

28. The computer of claim 26 further comprising

a display with a background responsive to an operating state of either said means for recognizing the speech input or said means for recognizing the handwritten or hand-drawn input.

29. The display of claim 28 further comprising

means for changing screen tint.

30. The display of claim 28 further comprising

means for displaying lines.

31. A television or video cassette recorder remote control for processing handwritten input or hand-drawn input and speech input comprising:

selecting a first operating mode and a second operating mode before processing handwritten, hand-drawn, or speech input;

means for recognizing the speech input as either a command or text for an electronic document in accordance with said first operating mode;

means for recognizing the handwritten or hand-drawn input as a command, text for an electronic document, or a graphic for an electronic document in accordance with said second operating mode; and

means for switching said first operating mode or said second operating mode to change how the computer recognizes said speech input or said handwritten or hand-drawn input.

32. The television or video cassette recorder remote control of claim 31, wherein said means for recognizing the handwritten or hand-drawn input includes either a three-dimensional wireless mouse or a video gesture recognizer.

33. An electronic whiteboard for processing handwritten input or hand-drawn input and speech input comprising:

selecting a first operating mode and a second operating mode before processing handwritten, hand-drawn, or speech input;

means for recognizing the speech input as either a command or text for an electronic document in accordance with said first operating mode;

means for recognizing the handwritten or hand-drawn input as a command, text for an electronic document, or a graphic for an electronic document in accordance with said second operating mode; and

means for switching said first operating mode or said second operating mode to change how the computer recognizes said speech input or said handwritten or hand-drawn input.

34. The electronic whiteboard of claim 33, wherein said means for recognizing the handwritten or hand-drawn input includes either a three-dimensional wireless mouse, a pen, a touch screen, or a video gesture recognizer.

35. A computer for processing handwritten input or hand-drawn input and speech input comprising:

means for recognizing handwritten or hand-drawn input;

means for recognizing speech input; and

a mechanical switch for switching among operating modes, wherein said operating modes include

a first mode wherein said means for recognizing speech input recognizes speech as text,

a second mode wherein said means for recognizing speech input recognizes speech as a command,

a third mode wherein said means for recognizing handwritten or hand-drawn input recognizes handwritten input as text,

a fourth mode wherein said means for recognizing handwritten or hand-drawn input recognizes hand-drawn input as a graphic, and

a fifth mode wherein said means for recognizing handwritten or hand-drawn input recognizes handwritten or hand-drawn input as a command.

36. A pen tablet computer for processing handwritten input or hand-drawn input and speech input comprising:

means for recognizing handwritten or hand-drawn input;

means for recognizing speech input; and

a mechanical switch for switching among operating modes, wherein said operating modes include

a first mode wherein said means for recognizing speech input recognizes speech as text,

a second mode wherein said means for recognizing speech input recognizes speech as a command,

a third mode wherein said means for recognizing handwritten or hand-drawn input recognizes handwritten input as text,

a fourth mode wherein said means for recognizing handwritten or hand-drawn input recognizes hand-drawn input as a graphic, and

a fifth mode wherein said means for recognizing handwritten or hand-drawn input recognizes handwritten or hand-drawn input as a command.

37. A palm computer for processing handwritten input or hand-drawn input and speech input comprising:

means for recognizing handwritten or hand-drawn input;

means for recognizing speech input; and

a mechanical switch for switching among operating modes, wherein said operating modes include

a first mode wherein said means for recognizing speech input recognizes speech as text,

a second mode wherein said means for recognizing speech input recognizes speech as a command,

a third mode wherein said means for recognizing handwritten or hand-drawn input recognizes handwritten input as text,

a fourth mode wherein said means for recognizing handwritten or handwritten input recognizes hand-drawn input as a graphic, and

a fifth mode wherein said means for recognizing handwritten or handwritten input recognizes handwritten or hand-drawn input as a command.

38. A telephone for processing handwritten input or hand-drawn input and speech input comprising:

means for recognizing handwritten or hand-drawn input;

means for recognizing speech input; and

a mechanical switch for switching among operating modes, wherein said operating modes include

a first mode wherein said means for recognizing speech input recognizes speech as text,

a second mode wherein said means for recognizing speech input recognizes speech as a command,

a third mode wherein said means for recognizing handwritten or hand-drawn input recognizes handwritten input as text,

a fourth mode wherein said means for recognizing handwritten or hand-drawn input recognizes hand-drawn input as a graphic, and

a fifth mode wherein said means for recognizing handwritten or hand-drawn input recognizes handwritten or hand-drawn input as a command.

39. A television or video cassette recorder remote control for processing handwritten input or hand-drawn input and speech input comprising:

means for recognizing handwritten or hand-drawn input;

means for recognizing speech input; and

a mechanical switch for switching among operating modes, wherein said operating modes include

a first mode wherein said means for recognizing speech input recognizes speech as text,

a second mode wherein said means for recognizing speech input recognizes speech as a command,

a third mode wherein said means for recognizing handwritten or hand-drawn input recognizes handwritten input as text,

a fourth mode wherein said means for recognizing handwritten or hand-drawn input recognizes hand-drawn input as a graphic, and

a fifth mode wherein said means for recognizing handwritten or hand-drawn input recognizes handwritten or hand-drawn input as a command.

40. An electronic whiteboard for processing handwritten input or hand-drawn input and speech input comprising:

means for recognizing handwritten or hand-drawn input;

means for recognizing speech input; and

a mechanical switch for switching among operating modes, wherein said operating modes include

a first mode wherein said means for recognizing speech input recognizes speech as text,

a second mode wherein said means for recognizing speech input recognizes speech as a command,

a third mode wherein said means for recognizing handwritten or hand-drawn input recognizes handwritten input as text,

a fourth mode wherein said means for recognizing handwritten or hand-drawn input recognizes hand-drawn input as a graphic, and

a fifth mode wherein said means for recognizing handwritten or hand-drawn input recognizes handwritten or hand-drawn input as a command.

* * * * *

UNITED STATES PATENT AND TRADEMARK OFFICE

# CERTIFICATE OF CORRECTION

PATENT NO. : 6,438,523 B1
DATED : August 20, 2002
INVENTOR(S) : John A. Oberteuffer et al.



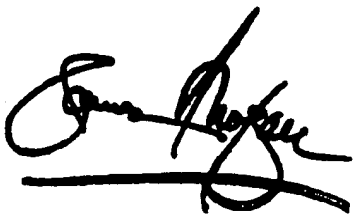
It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 8,
Line 50, after "handwritten" insert a comma.

Column 15,
Line 7, "handwritten input" should read -- hand drawn input --.
Line 10, "handwritten input" should read -- hand-drawn input --.

Signed and Sealed this

Fourth Day of February, 2003



JAMES E. ROGAN
*Director of the United States Patent and Trademark Office*

US005914707A

# United States Patent [19]

Kono

[11] Patent Number: 5,914,707

[45] Date of Patent: *Jun. 22, 1999

[54] **COMPACT PORTABLE AUDIO/DISPLAY ELECTRONIC APPARATUS WITH INTERACTIVE INQUIRABLE AND INQUISITORIAL INTERFACING**

[75] Inventor: **Mitsuru Kono**, Suwa, Japan

[73] Assignee: **Seiko Epson Corporation**, Tokyo, Japan

[ * ] Notice: This patent issued on a continued prosecution application filed under 37 CFR 1.53(d), and is subject to the twenty year patent term provisions of 35 U.S.C. 154(a)(2).
This patent is subject to a terminal disclaimer.

[21] Appl. No.: **08/455,457**

[22] Filed: **May 31, 1995**

**Related U.S. Application Data**

[63] Continuation of application No. 07/890,350, May 22, 1992, which is a continuation of application No. 07/496,788, Mar. 21, 1990, abandoned.

[30] **Foreign Application Priority Data**

Mar. 22, 1989 [JP] Japan ..... 1-70256
Oct. 12, 1989 [JP] Japan ..... 1-265679

[51] **Int. Cl.**[6] ..... **G09G 5/00**

[52] **U.S. Cl.** ..... **345/173**; 345/901

[58] **Field of Search** ..... 345/156, 157, 345/173, 179, 169, 901, 902, 348, 349, 352, 354, 356, 357, 978; 364/705.06, 705.02, 705.03, 708.01; 434/178, 179, 307, 308, 322, 323

[56] **References Cited**

U.S. PATENT DOCUMENTS

| | | | |
|---|---|---|---|
| D. 277,962 | 3/1985 | Thom . | |
| 4,159,417 | 6/1979 | Rubincam . | |
| 4,545,023 | 10/1985 | Mizzi . | |
| 4,633,323 | 12/1986 | Haberkern et al. . | |
| 4,634,845 | 1/1987 | Hale et al. . | |
| 4,639,225 | 1/1987 | Washizuka . | |
| 4,664,299 | 5/1987 | Krenz . | |
| 4,667,299 | 5/1987 | Dunn . | |
| 4,779,080 | 3/1986 | Coughlin et al. . | |
| 4,796,100 | 1/1989 | Sakaguchi . | |
| 4,849,827 | 7/1989 | Hashimoto et al. . | |
| 4,855,725 | 8/1989 | Fernandez | 345/173 |
| 4,868,653 | 9/1989 | Golin et al. . | |
| 4,868,764 | 9/1989 | Richards . | |
| 4,873,764 | 10/1989 | Ishikawa et al. . | |
| 5,031,119 | 7/1991 | Dulaney et al. | 345/901 |
| 5,379,057 | 1/1995 | Clough et al. | 345/173 |

FOREIGN PATENT DOCUMENTS

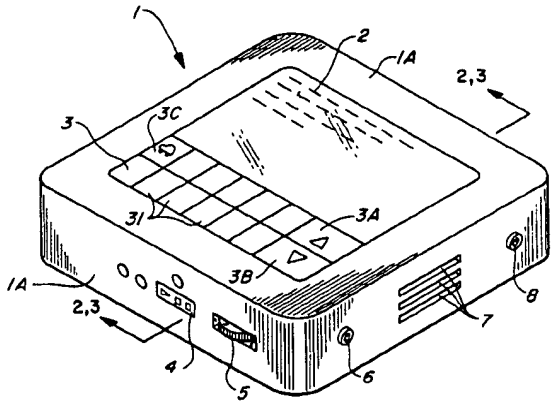
| | | |
|---|---|---|
| 0149101 | 7/1985 | European Pat. Off. . |
| 0193996 | 9/1986 | European Pat. Off. . |
| 0252646 | 1/1988 | European Pat. Off. . |
| 0278051 | 8/1988 | European Pat. Off. . |
| 56 073985 | 6/1981 | Japan . |
| 62-0279585 | 3/1987 | Japan . |
| 62 080837 | 4/1987 | Japan . |
| 2-52588 | 11/1987 | Japan . |
| 63-157890 | 10/1988 | Japan . |

*Primary Examiner*—Chanh Nguyen

[57] **ABSTRACT**

A compact size, portable audio/display electronic apparatus having a compact mass memory with the capability of entry of user originated inquirable and inquisitorial input relative to currently displayed information and the random access capability relative to the compact mass memory in retrieving data responsive to such input. The apparatus includes a recording and reproducing device for recording and reproducing data in the form of alphanumeric information, diagram graphics, animation graphics, music, voice and other audio reproducible material. The reproducing device is employed only for the reproduction of the data, retrieving means for only retrieving the data for reproduction on a random access basis. Further, the display and reproducing device is arranged in one compact housing for the synchronized reproduction of at least two kinds of such data, e.g., the display of information and the audio reproduction of information directly associated with the displayed information.

**3 Claims, 7 Drawing Sheets**



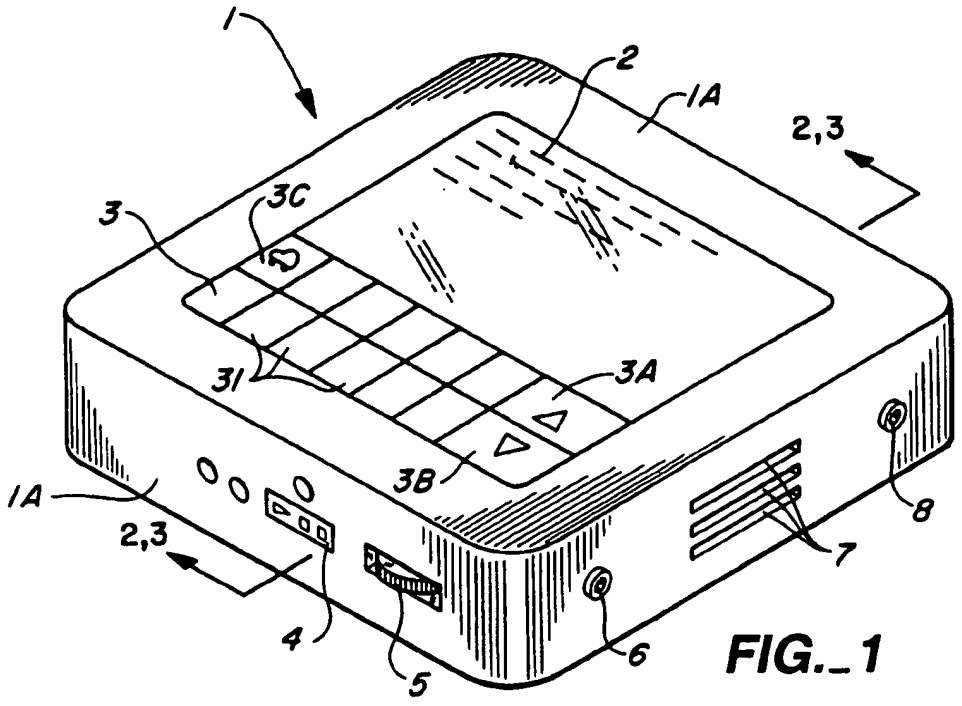



FIG._1

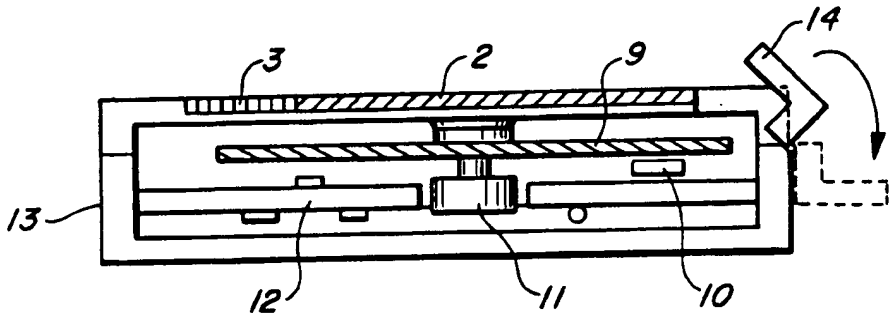



FIG._2

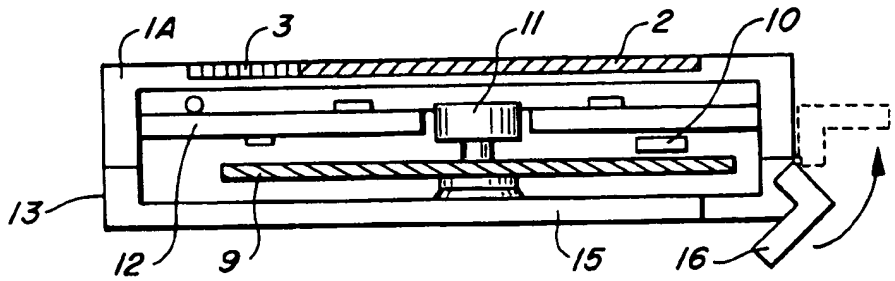



FIG._3

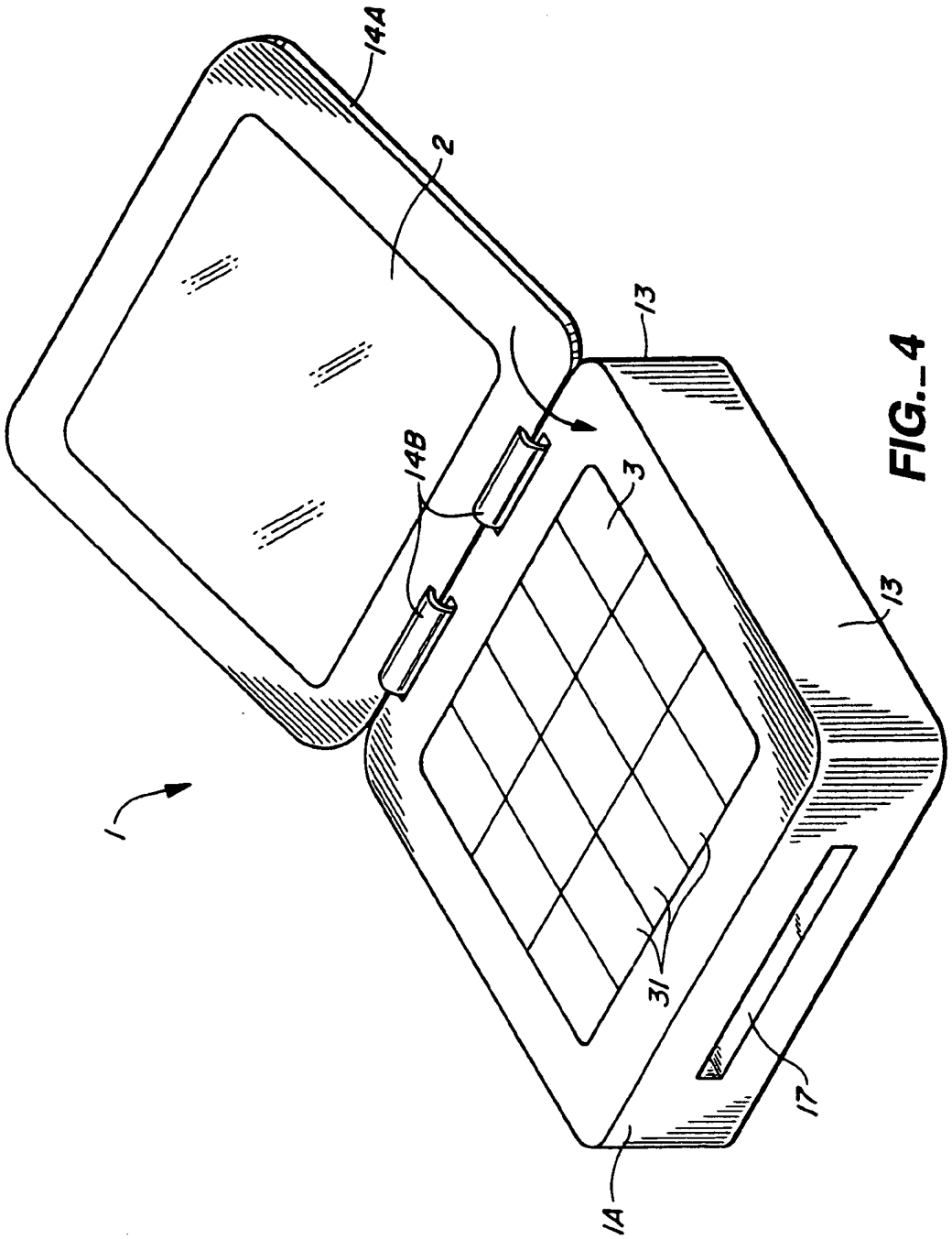



FIG._4

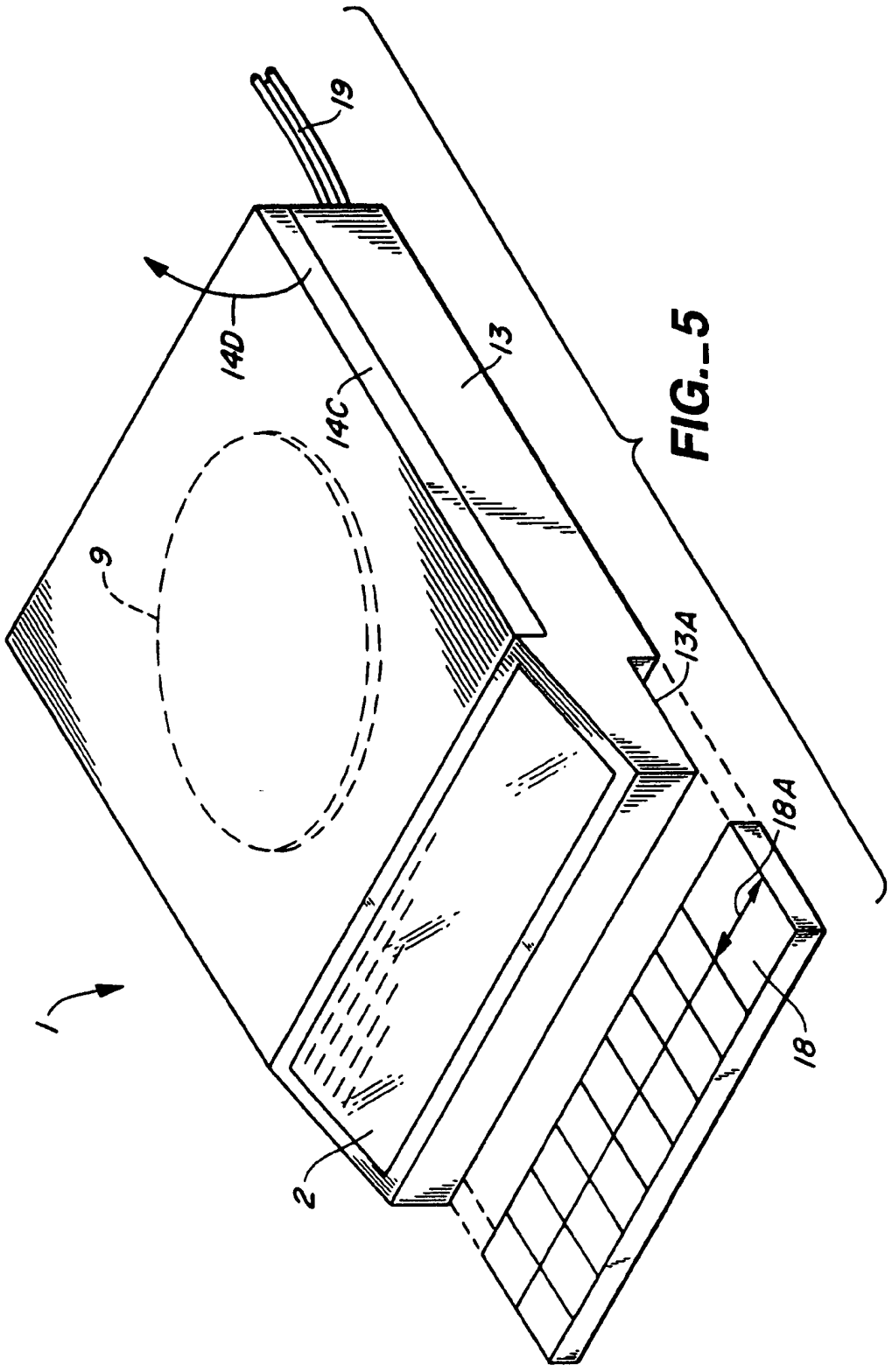



FIG._5

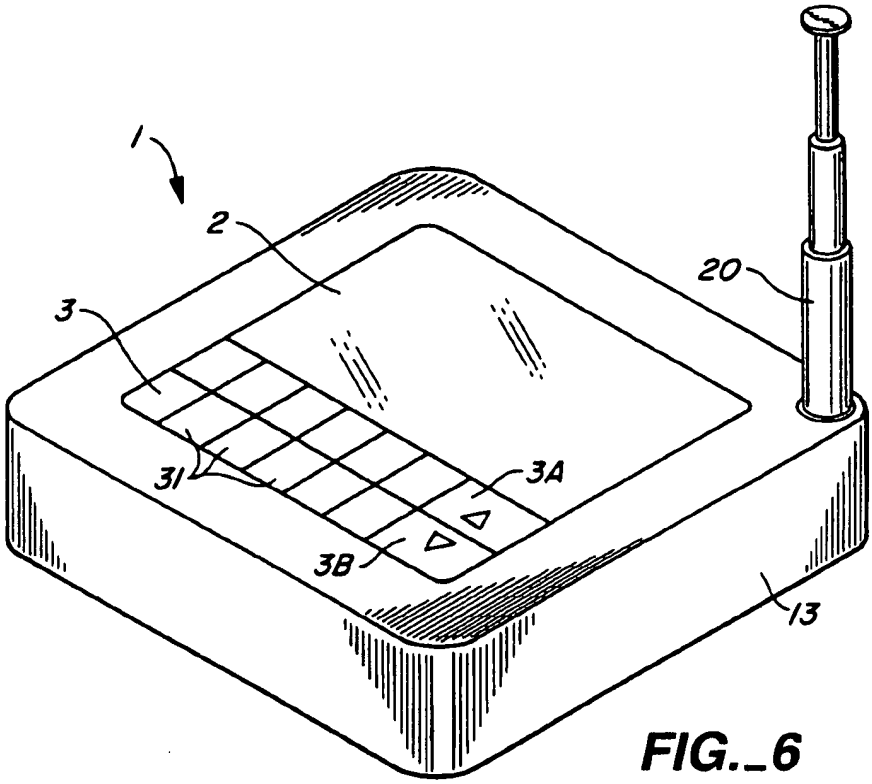



FIG._6

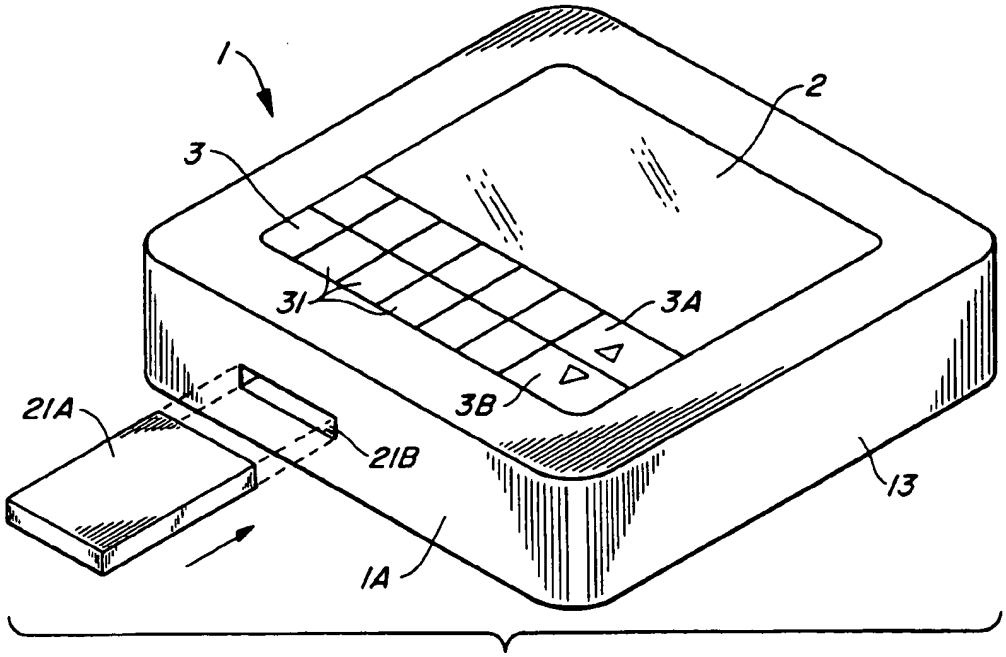



FIG._7

FIG._8A

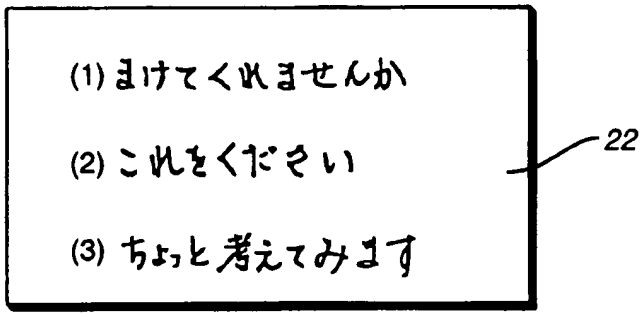



FIG._8B

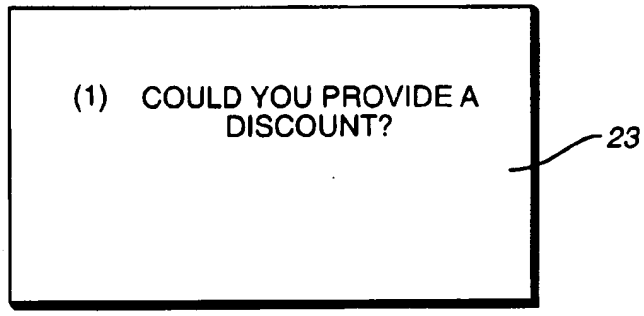
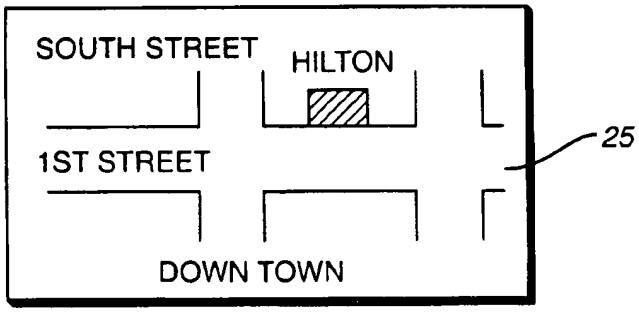



FIG._9A


FIG._9B



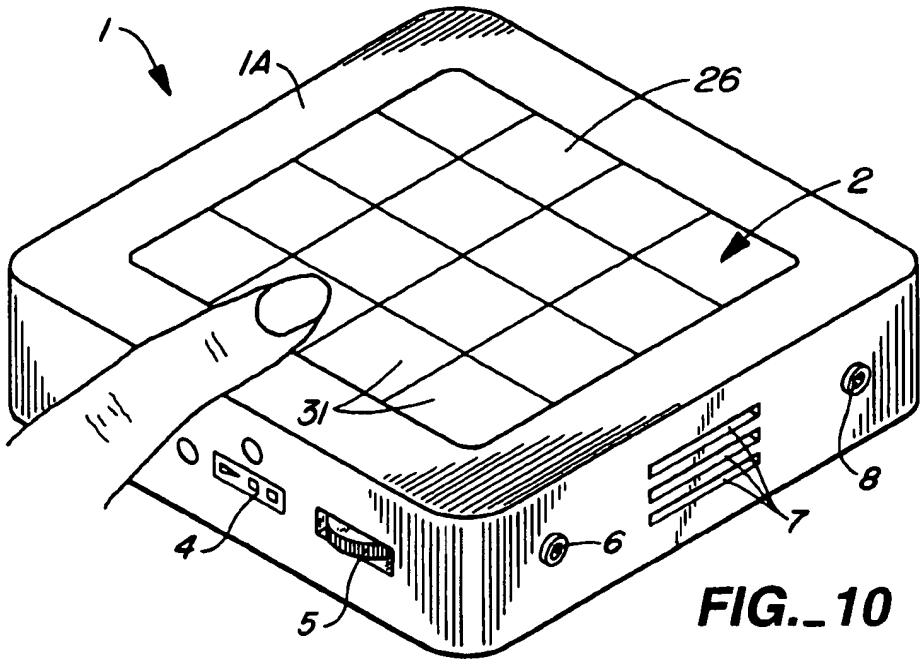



FIG._10

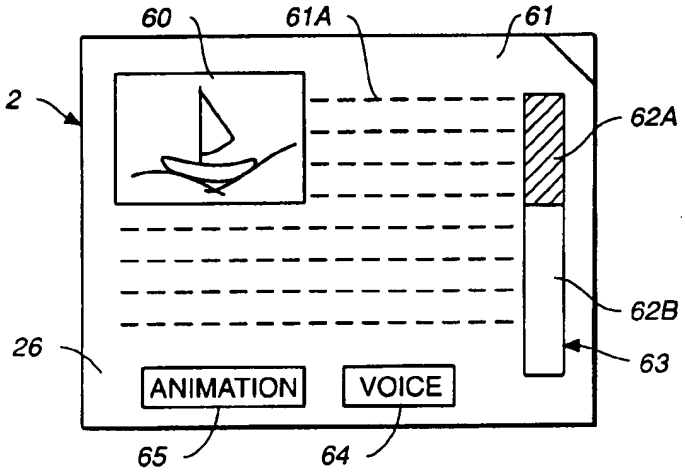



FIG._12

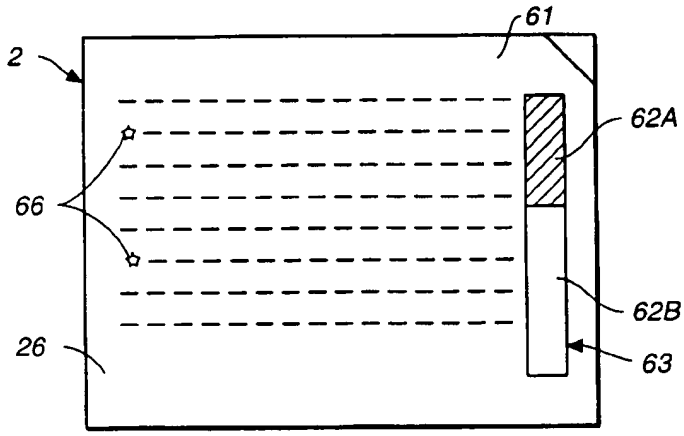



FIG._13

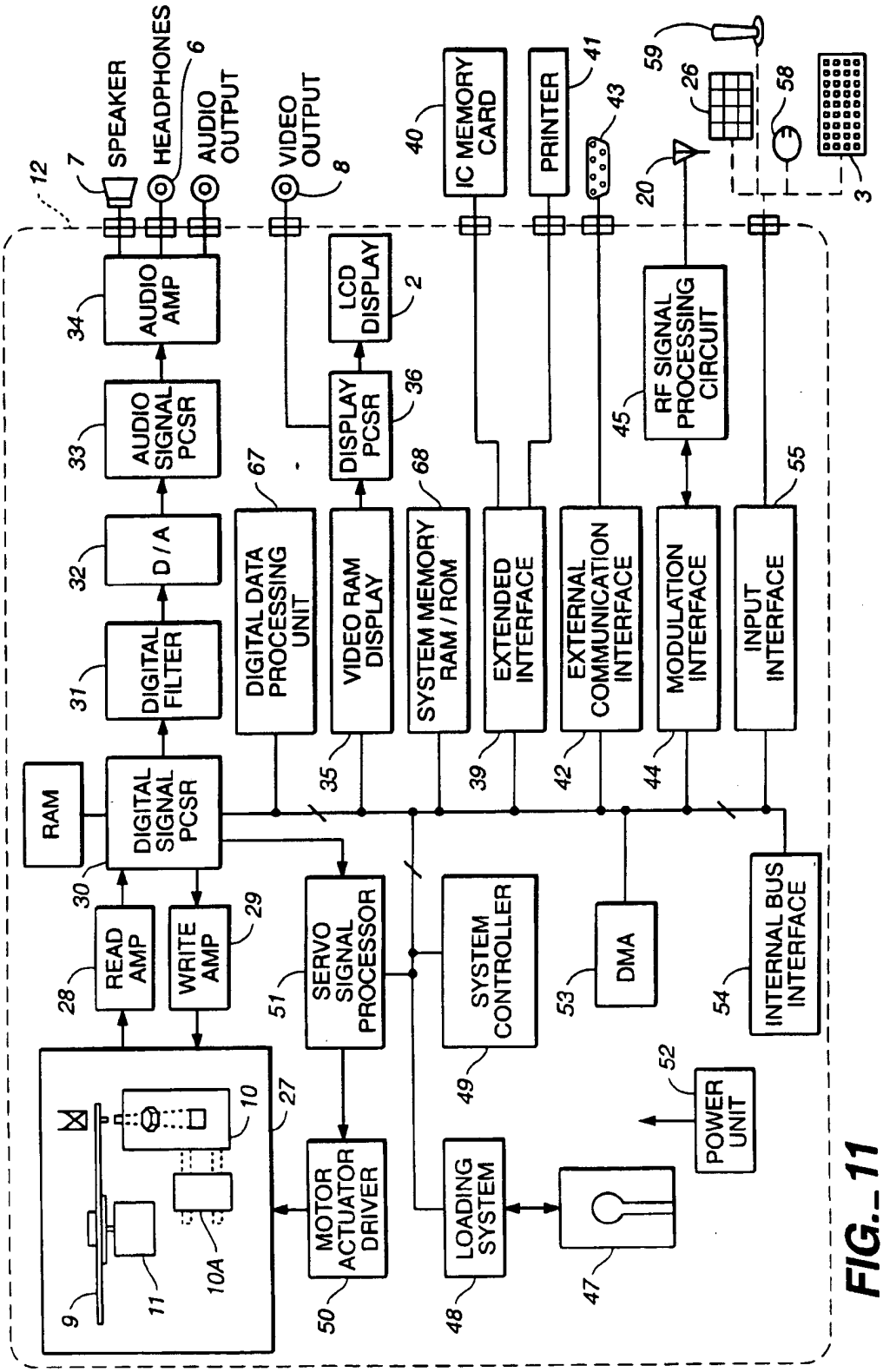

SPEAKER
HEADPHONES
AUDIO OUTPUT
VIDEO OUTPUT
IC MEMORY CARD
PRINTER
AUDIO AMP
AUDIO SIGNAL PCSR
D / A
DIGITAL FILTER
DIGITAL SIGNAL PCSR
RAM
READ AMP
WRITE AMP
LCD DISPLAY
DISPLAY PCSR
DIGITAL DATA PROCESSING UNIT
VIDEO RAM DISPLAY
SYSTEM MEMORY RAM / ROM
EXTENDED INTERFACE
EXTERNAL COMMUNICATION INTERFACE
MODULATION INTERFACE
INPUT INTERFACE
RF SIGNAL PROCESSING CIRCUIT
SERVO SIGNAL PROCESSOR
SYSTEM CONTROLLER
DMA
INTERNAL BUS INTERFACE
MOTOR ACTUATOR DRIVER
LOADING SYSTEM
POWER UNIT
12
7
6
8
40
41
43
59
26
58
20
3
34
33
32
31
67
2
36
68
45
55
35
39
42
44
30
28
29
51
49
53
54
52
27
10
10A
11
9
50
48
47

FIG._11

# COMPACT PORTABLE AUDIO/DISPLAY ELECTRONIC APPARATUS WITH INTERACTIVE INQUIRABLE AND INQUISITORIAL INTERFACING

This is a continuation of U.S. patent application Ser. No. 07/890,350, filed May 22, 1992 which is a continuation of U.S. patent application Ser. No. 07/496,788 filed Mar. 21, 1990, now abandoned.

## BACKGROUND OF THE INVENTION

This invention relates generally to a compact and portable audio/display device for displaying and reproducing data in form of alplanumeric information, diagrams, graphics, music, and sound which are recorded on a mass storage medium, such as, an optical medium, a magneto-optical recording medium, CD, CD-ROM or the like, wherein at least two kinds of data are randomly accessed for synchronized and uniform display and audio reproduction with the provision of a mass storage medium in a portable apparatus having an interactive user interface providing for inquirable and inquisitorial input by the user to random access and retrieve demanded response based upon information stored in such a memory. By "compact", it is meant that the apparatus is self contained, i.e., it contains all the functional components of the apparatus in a unitary handheld size unit. By "portable", it is meant that the apparatus can be comfortably held by the user in one hand and operated by the user with the other hand, contains its own power source and is easily carried from one place to another, for operation independent of any other kind of information source.

Compact and portable type personal data audio/display apparatus are now in common use, for example, compact electronic pocketbooks, notebooks or organizers which include various functional combinations, for example, one or more of the following functions: note pad, diary, phone directory, calendar, world clock, as well as compact electronic dictionaries and electronic translators. However, all of these types of compact and portable type personal data display apparatus are generally limited to alphanumeric and symbol data input/output. For example, a compact microcomputer notebook or pocket organizer is useful to collect dictated communications, such as, notes relative to an overseas trip and the user may visually read written words or sentences that appear on the organizer's display, but for such a small slim device there is no facility to orally reproduce the voice pronunciation of the words and sentences due to its compactness. To include concurrent audio reproduction capability and user inquirable and inquisitorial interfacing while retaining their compact size is considered not practical under present technology for such organizers.

Other compact and portable apparatus provide for concurrent audio speech/display output. One such application is disclosed in Japanese Laid Open publication No. 62-279585 published on Dec. 4, 1987. This publication discloses a portable CD and CD-ROM disc player apparatus which includes concurrent display of alphanumeric and symbol information with audio output, such as speech, associated with the displayed information. Both the audio output and visually displayed data is stored on the CD-ROM. Input means to the compact player apparatus provides interactive functions for stop, fast forward, fast return and start playback command modes. However, this type of interactiveness is of limited capability, i.e., there is no provision of providing a sufficiently large mass memory capability in combination with an input means that permits a user to actively



interact with data concurrently displayed and audibly presented to the user, i.e., permitting inquirable and inquisitorial interfacing with data stored in and display on the apparatus for the purpose of invoking a response thereto, such as is common practice with, for example, non-handheld personal computers (PC), workstations, etc. This is primarily due to the need for compact mass storage memory and the necessary electronics required to provide random access to the mass storage memory to locate, retrieve, decode and display and audibly present data in response to user inquiries.

Another example of such a portable audio/visual electronic apparatus is disclosed in U.S. Pat. No. 4,639,225. The portable apparatus utilizes an insertable ROM memory card which contains both audio and video information. The memory card is inserted into the device for display of information on the apparatus display in conjunction with reproduction of speech and other audio information, for example, such as the display and audio reproduction of novels, cartoons and dictionary terms. However, the ability of a portable video/audio electronic apparatus of this type to provide good graphic display quality or generate speech and other audio reproduction via an IC memory card circuit is generally of bad quality. Furthermore, there is no means provided by which a user may actively interact with the data presented on the insertable memory, such as inquirable and inquisitorial interfacing by the user relative to current displayed data. Again, user interfacing in such compact and portable apparatus is aimed at a low level of controlling the presentation of the data, i.e., interfacing is limited to handling of the display and reproduction of audio information, such as start, stop, and sequential replay, either forward or reverse, of stored data, without the capability of entertaining inquiries by the user and providing information in the form of a response or answer to a presented inquiry.

Thus, the disadvantage of employing an IC memory or IC memory card is that the memory capacity for graphics and audio reproduction is too limited and the sound quality reproduction is not acceptable. Moreover, the same is true with the use of a tape memory for sound which is undesirable because random access is not efficiently possible and takes too long to retrieve and reproduce the selected sound portion. The use of a tape medium for recording takes a comparatively long period of time for random access of selected data or information desired for playback. Further, as previously indicated, a mass memory, such as a CD-ROM, may be easily employed as an external memory device for a PC, but these types of memories are inconvenient and expensive to consider for application apart from a personal computer application, such as for an application in a compact electronic organizer or other small portable electronic apparatus. An example of the foregoing is disclosed in U.S. Pat. No. 4,855,725 involving a portable electronic apparatus comprising an electronic simulated book that displays on a LCD panel a page of a book from a series of limited pages available in a limited memory provided in the portable apparatus. The apparatus provides for user input to display a selected page from a starting page to an ending page stored in the limited memory and also to consecutively display pages in ordered sequence set in memory. However, there is no ability to randomly select pages or input inquires into the apparatus based upon information displayed. Furthermore, since there is no mass storage capabilities in this portable apparatus, the user is required to be in close proximity for wireless communication to a PC which has peripheral mass storage of all the book pages on a CD-ROM. Thus, the limited supply of pages stored in the portable apparatus are replenished from the remote PC. As a result, there is no

ability to employ the portable apparatus for long periods of time at a remote location from the PC so that any true portability of the apparatus is limited by its proximity to a nonportable PC unit. In this sense, the display apparatus is not truly portable.

Furthermore, there is no means in the portable apparatus of U.S. Pat. No. 4,855,725 for the user to actively interact with the apparatus at a level of entering queries into the apparatus pertaining to concurrently displayed information in order to solicit audio responses and/or responsive displayed illustrations to satisfy the entered query. Rather, as in the case of previously discussed Japanese Laid Open publication No. 62-279585, there is conventional input means for the purpose of stopping, starting, a sequencing information displayed and audibly reproduced without further higher level and more complex user interactiveness relative to the stored information. In this connection, while publication No. 62-279585 provides for a mass memory in the form of a CD or a CD-ROM, the function of the CD-ROM is to hold programs for their selective transfer to the memory of the apparatus for execution by the apparatus CPU without any capability of user inquirable and inquisitorial interactiveness relative to data stored on the CD-ROM.

Furthermore, it is difficult to input data to CD-ROM memories by means of an I/O device, such as, a compact keyboard. Thus, a compact portable apparatus that provides complement keyboard input for inputting information are small and not easily useable due to their limited size for the sake of compactness. Alternatively, a smaller key input means to function as a user interface could provide desired inquirable and inquisitorial interfacing with data stored in, displayed on and audibly reproduced by the apparatus.

It is an object of this invention to provide a solution to the foregoing mentioned problems.

It is another object of this invention to provide inquirable and inquisitorial interfacing in a compact portable audio/display electronic apparatus.

It is another object of this invention to provide a compact and portable audio/display electronic apparatus that provides for a large mass memory capability and for a high level of interactiveness between the user and stored data in the apparatus capable of providing concurrent visual and audio data in response to inquiries provided by a user relative to displayed data.

It is another object of the present invention to provide a compact display and reproducing apparatus which can immediately retrieve a substantial amount of data or information recorded on small optical media by randomly accessing such data or information and, further, synchronizes the reproduction of at least two kinds of such data, e.g., the display of information and the audio reproduction of information directly associated with the displayed information.

## SUMMARY OF THE INVENTION

According to this invention, there is provided in a compact size, portable audio/display electronic apparatus having a compact mass memory with the capability of entry of user originated inquirable and inquisitorial input relative to currently displayed information and the random access capability relative to the compact mass memory in retrieving data responsive to such input. By "inquirable", it is meant the capability of permitting the user relative to a particular application to inquire into a subject matter provided in the compact mass memory or subject a subject recorded in the mass memory to inquiry. By "inquisitorial", it is meant the capability of permitting a user relative to a particular application to be inquisitory, to search and examine data or portions of audibly reproduced/displayed data and interact with the displayed data via iconic representations to obtain additional displayed or audibly reproduced data, e.g., graphic animation, providing additional explanation and enhancement concerning displayed data, e.g., text.

The compact electronic data display apparatus of this invention comprises recording and reproducing means for recording and reproducing data in the form of alphanumeric information, diagram graphics, animation graphics, music, voice and other audio reproducible material. The reproducing means is employed only for the reproduction of the data, retrieving means for only retrieving the data for reproduction on a random access basis. Further, the display and reproducing means is arranged in one compact housing for the synchronized reproduction of at least two kinds of such data, e.g., the display of information and the audio reproduction of information directly associated with the displayed information.

By arranging the data display and data reproducing means in one compact housing, it is possible with this invention to further miniaturize the entire apparatus. Further, the employment of an optical medium relative to this invention renders it possible to display a large amount of data in the form of alphanumerics, simple diagram and animation graphics and to reproduce audio data, such as sound, voice and music, being synchronized in presentation with the displayed data. The present invention provides a simple, compact and portable electronic apparatus for recording visual and audio data, randomly retrieving such data from the optical memory and synchronizing the displayed data with the audio reproduced data. Also, application software may be provided in the optical memory or in the form of an IC memory in combination with the optical memory to perform various functions on demand.

Other objects and attainments together with a fuller understanding of the invention will become apparent and appreciated by referring to the following description and claims taken in conjunction with the accompanying drawings.

## BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1** is an isometric view of an external configuration for the compact portable audio/display electronic apparatus of this invention including iconic input interfacing and associated display and audio output.

FIG. **2** is a cross sectional side elevation of FIG. **1** illustrating a first embodiment relative mass memory access.

FIG. **3** is a cross sectional side elevation of FIG. **1** illustrating a second embodiment relative mass memory access.

FIGS. **4**, **5**, **6**, and **7** are respectively isometric views of four other external configurations for the compact portable audio/display electronic apparatus of this invention.

FIGS. **8A**, **8B**, **9A** and **9B** are examples of illustrating an application of the interactive interfacing provided by this invention relative to the synchronized displayed data and reproduced audio via random access of data stored relative to mass storage memory.

FIG. **10** is an isometric view of a further embodiment for the compact portable audio/display electronic apparatus of this invention including a combination display and overlying transparent input key panel.

FIG. **11** is a diagram of the control circuit employed relative to the several configurations of the compact portable audio/display electronic apparatus comprising this invention.

FIGS. 12 and 13 are examples illustrating an application of the interactive interfacing provided by this invention relative to an electronic book.

## DESCRIPTION OF THE PREFERRED EMBODIMENTS

Reference is now made to FIG. 1 which illustrates an external view of one embodiment of a compact portable audio/display electronic apparatus 1 comprising this invention. Apparatus 1 includes casing 13 having display 2 on its top surface and an input keyboard 3 for operating display 2. Display 2 is preferably a liquid crystal display because of its low power requirement for the display of data. Any state-of-the-art liquid crystal display can be utilized, for example, a monochrome or multi-colored or full colored display.

On the front side surface 1A of casing 13 are operating buttons 4 for turning on apparatus 1 and for accessing the drive for a mass memory medium 9, such as, a CD, CD-ROM, or other such optical memory medium and a volume control 5 for controlling the volume of the audio output at speaker output 7. Side face 1B of casing 13 includes a headphone output terminal 6, speaker output 7 and CRT output terminal 8.

As shown in FIG. 2, cap portion 14 is rotatably mounted relative the top surface of casing 13 adjacent to LCD display 2. Cap portion 14 may be released by an operating button 4 and lifted to its open position to permit the insertion or removal of optical medium 9 from within the interior of casing 13, which medium is positioned on a spindle of spindle motor 11. Thus, cap portion 14 provides for easy insertion and removal of medium 9 relative to casing 13. In FIG. 3, another embodiment for insertion and removal of optical medium 9 is shown. In this case, cap portion 16 is rotatably mounted relative to the bottom side 15 of casing 13, i.e., on the side opposite to display 2. Cap portion 16 may be lifted to its open position to permit the insertion or removal of optical medium 9 from within the interior of casing 13 positioned on a spindle of spindle motor 11. Thus, cap portion 16 provides for easy insertion and removal of medium 9 relative to casing 13 while being positioned in a region on casing 13 separate from the region of display 2 so that the structure of display 2 is not stressed by pressure applied relative to the opening and closure of cap portion 16.

In FIGS. 2 and 3, schematically shown is a circuit board as supported in casing 13 from impact and shock and upon which is the IC circuit for control circuit system 12, illustrated in FIG. 11. System 12 controls the operation of medium 11 and the recording and retrieving of data relative to medium 11 via optical pickup 10 for display and audio reproduction.

In connection with both embodiments in FIGS. 2 and 3, the size of optical medium 9, e.g., a CD-ROM, is preferably small such as only 8 cm in diameter, which is sufficiently small in size for compact portable apparatus 1 so as to be easily carried in one's hand. However, it is also possible to use a larger standard size CD, e.g., 12 cm, in order to support a larger display 2 thereby rendering it possible to display not only alphanumerics but also to display larger area graphics, such as a large map or the like.

Input keyboard 3 may be a transparent touch panel or keyboard comprising individual keys 3' and also includes a card insertable card beneath the transparent top of keyboard 3 for making designated changes to the nomenclature or symbols employed in conjunction with the key caps of keyboard 3. Keyboard 3 need not be a full complement alphanumeric keyboard and may consist, for example, of several keys 3'. Fourteen such keys 3' are shown in FIG. 1. Two keys, 3A and 3B, are for scrolling to a previous display or to the next display of previously selected data relative to keys 3'. Keyboard 3 is employed to provide, for example, a two step interactively inquirable interfacing by the user by permitting the user to initially select a prerecorded subject group prerecorded on medium 9, display the subject group and thereafter select a particular item in that group for further treatment and display. Therefore, it is possible, for example, with keyboard 3, to first choose a subject group by selecting an appropriate symbol marked key 3' followed by selecting a particular item in the displayed subject group for further detailed display by selecting another symbol marked key 3'. Alternatively, a second level selection (i.e., an item selected from a retrieved subject group) may be a reassignment of function of the keys 3' on keyboard 3 to perform different tasks as compared to an assigned function of the same keys at the first level of selection (i.e., selection of the particular subject group).

FIG. 8 illustrates an example of inquirable interfacing via an application of compact portable audio/display electronic apparatus 1 employed as a pocket notebook employed, for example, in recording data or information relative to an overseas trip. In FIG. 8A, three Japanese sentences at 22 have been previously recorded on medium 9 concerning the trip and are prompted to be visually displayed as well audibly reproduced as speech via selection of a symbol or numerically identified key 3' on keyboard 3, such as a particular key or key combination indicative of the search location or functioning as a pointer in memory for the location of the subject group for purposes of retrieving and decoding of the selected digital data. One Japanese sentence, item or sentence (1), is then chosen by the user for retrieval relative to the subject group of sentences (1), (2), and (3) by selection of another symbol or numerically identified key 3' on keyboard 3. Then, the chosen sentence (1) is translated by apparatus 1 into English, as indicated in FIG. 8B, and thereafter the English translated sentence (1) is displayed on display 2 as indicated at 23 along with audibly reproduced data in the form of speech vocalizing the English sentence with good audio quality at speaker 7.

FIG. 9 is another example of an inquirable interfacing wherein a directory has been provided in apparatus 1 on medium 9 to include a list of hotels for a particular city, and this subject group may be selected via a particular key 3', such as, key 3C in FIG. 1, designated as a pointer memory for one of a particular subject groups. Then, the list of hotels, illustrated at 24 in FIG. 9A, is displayed on display 2. At this point, the function assignment of keys 3' are changed relative to the selection of an item from the displayed subject group. The user then selects a particular item, i.e., a hotel by selecting a numerically identified key 3' on keyboard 3 indicative of that selection, in this case, the hotel indicated as hotel (1) from the list 24. The graphic location of this hotel is thereafter displayed as a simple map 25 on display 2, which map is stored on medium 9 and is accompanied with reproduced audio explaining the location of the hotel, for example, on First Street relative to corresponding cross streets in the city. Thus, optical medium 9 provides for a compact medium for storing a comparatively large amount of data or information which can be displayed on apparatus display 2 accompanied with necessary translation or graphic diagrams concurrently reproduced with sound or music whereas conventional compact and portable apparatus of this kind provide only for the display of alphanumeric type data.

FIGS. 12 and 13 illustrate an example of inquisitional interfacing relative to apparatus 1 relative to an electronic

book comprising a series of electronic pages recorded on medium **9**. Here, individual pages of text of the book, such as a page 61A illustrated at **61** in FIG. **12**, may be displayed on display **2** and for inquisitional embellishment of the text for enlightenment of the user is provided by employing a transparent coordinate touch panel **26** in FIG. **10** to provide a series of displayed icons, **63**, **64** and **65**, that are displayed with the displayed text **61A** and appear to the user through the transparent touch panel **26**. The activation of these displayed icons permits the user to receive additional relevant data related to currently displayed data, such as text **61A**, to enhance the interest and inquisitiveness of the user as well as provide supplemental data to the user to function as a tutorial and enhance the knowledge and further understanding of the user with respect to the subject of the displayed text **61A**. The icon bar indicator **63** shows the relationship of page **61** relative to entire electronic book of pages as represented by the length of bar **63** so that the user has a visual concept of the number pages in the entire book. The hatched portion **62A** of indicator bar **63** is a visual indication of the number the pages that proceed the display of page **61** and portion **62B** of indicator bar **63** is a visual indication of the number of pages that come after page **61**. The display of page **61** includes one stationary graphic image **60** along with text **61A**. Simple graphic animation of image **60** can be displayed by selecting displayed icon **65** to cause the stationary image **60** to move in animation such as through two or move motion frames, e.g., an animation of a sail boat moving over the waves on a body of water. Audio speech may accompany the display of page **61** by selecting voice icon **64**. For example, reproduced audio speech may footnote or exemplify the displayed text. It is possible to accelerate the electronic advancement and display of pages by more than several pages at one time in conjunction with bar **63** by the user finger pressing on transparent coordinate positions along the length of iconic indicator bar **63** to cause selection and retrieval of a different book page either as a previous page within the confines of bar portion **62A** or future page within the confines of bar portion **62B**.

In FIG. **13**, data is written into mass memory medium **9**, such as, a magneto-optical memory of apparatus **1** and pointers **66** may be recorded in conjunction with recorded data, e.g., text **61A**, that is considered particularly important in nature relative to a particular data section, for example. Pointers **66** are also retained in memory and corresponding encoded pointers to this data in memory is employed for retrieving at a later time and displayed for viewing, as is known in the art. Such marked data may also be printed via a connected printer as explained in connection with the embodiment of FIG. **5**, discussed later. Another application of pointer marked data is to provide the function of selectively enlarging displayed alphanumeric information which enables users with poor eye sight to more easily read such information.

FIGS. **4** through **7** disclose other types of exterior configurations that may be employed relative to compact portable audio/display electronic apparatus **1**. In FIG. **4**, display **2** is formed as part of a cap portion **14A** over casing **13** and cap portion **14A** is rotatably attached at **14B** to the upper portion of casing **13**. Further, optical medium **9** is inserted in a slot **17** provided in the front side face **1A** of casing **13**.

In FIG. **5**, display **2** is arranged on casing **13** separately from a cap portion **14C** and cap portion **14C** is rotatably attached to the upper portion of casing **13**. Optical medium **9** is inserted into apparatus **1** by opening cap portion **14C**, as indicated by arrow **14D**, and placing medium **9** on the spindle support positioned beneath cap portion **14C**. In this embodiment, input keyboard **18** is removably attached to the forward underside of casing **13** in a recess **13A**. Keyboard **18** may be attached to the back side of recess **13A** by first positioning keyboard **18** into recess **13A** and then moving it laterally as indicated by arrow **18A** to bring about its coupling to casing **13**. Keyboard **18** may be connected as an input to system **12** by means of a cable (not shown) or by means of IR transceivers as is known in the art, e.g., illustrated in U.S. Pat. No. 4,855,725, previously discussed. Also, data stored on medium **9** in apparatus **1** and displayed on display **2** may also be printed out with a printer attached to connecting cable **19**. Also, apparatus **1** may be employed as an external memory device for I/O connection to other office automation apparatus, e.g., personal computers, word processors or the like, by connecting such office apparatus to apparatus **1** via connecting cable **19**.

FIG. **6** is an embodiment of apparatus **1** wherein it is possible to display visual data concurrently with audio data via receipt of a transmitted signal at antenna **20** to a transmitter/receiver in system **12** of apparatus **1**. As an example, in a zoo, a user of apparatus **1** may approach a caged zoo animal and also come within the transmission range of a transmitted signal at the animal cage receivable at antenna **20**. The received signal provides both displayed and graphic data relating to the particular animal as well as an audio explanation that accompanies the displayed data as well as other data pertaining to the particular animal.

FIG. **7** is an embodiment of apparatus **1** which enables recording and reproducing of data via the employment of a RAM IC memory card **21A** insertable into slot **21B** in front face **1A** as well as an optical medium **9**, such as CD-ROM, insertable into apparatus **1** in the manner explained previously in connection with the embodiments of FIGS. **2** and **3**. Also, it is possible to record new data on optical medium **9**, employing either an erasable type magneto-optic medium or a write once type medium, supplied from an external database via, for example, a modem.

FIG. **10** shows another embodiment of the electronic apparatus wherein LCD display **2** comprises an overlying, transparent touch key panel **26**. When a portion of the coordinate detection surface representing keys **3'** is depressed, as indicated in FIG. **10**, the coordinate value of that portion is detected and is representative of the particularly selected key **3'**.

FIG. **11** is a circuit diagram of the control circuit system **12** employed with respect to the embodiments of apparatus **1** shown in FIGS. **1**, **4**, **5**, **6**, **7** and **10** and relative to the previously explained interactive inquirable and inquisitional interfacing discussed in connection with FIGS. **8**, **9**, **12** and **13**.

In FIG. **11**, compact mass storage record and playback unit **27** is connected in system **12** to digital signal processor **30** through READ amplifier **28** and WRITE amplifier **29** and motor/actuator driver **50** is connected to optical pickup actuator **10A** and spindle motor **11** in unit **27**. The heart of circuit system **12** is system controller **49** followed by digital signal processor **30**. System controller **49** contains the CPU and controller functions for executing instructions from system memory **68**, which includes both RAM and ROM memory. Functional operation of controller **49** is well known in the art, and processing input signals in accordance with a program stored in the ROM to control other processing circuits to be noted later and controlling the access, time sharing and synchronization of other circuit and I/O components connected in system **12** to data bus **70**. For the purpose of simplification, address buses and addressing

functions necessary to the operation of system controller **49** and addressed components connected to bus **70** are not shown but are well known in the art. Internal bus interface **54** controls access of connected components to bus **70** under the control of system controller **49**. DMA provides for direct peripheral to memory communication and control as is known in the art.

Connected to bus **70** is servo processor **51** which is connected to servo driver **50** for controlling the movement and operation of spindle motor **11** and pickup actuator **10A**. Also, digital signal processor **30** is connected to bus **70** and to servo signal processor **51** for the purpose of accessing recorded data or recording data on medium **9** under the control of system controller **49**.

The function of digital signal processor unit **30** is interleaving processing and to code data to be recorded on medium **9** via WRITE amplifier **29** and decode data retrieved from medium **9** via READ amplifier **28**. READ amplifier is all that is required relative to the employment of mass memory medium **9** as a CD-ROM. Video data compression is known in the art and is exemplified in U.S. Pat. Nos. 4,868,764 and 4,868,653. Digital signal processor **30** retrieves audio data, such as, voice or music, as well as display data and separates the data as coded and decodes, demodulates and decompresses the separated data. Audio data pass through digital filter **31**, D/A converter **32**, audio signal processing unit **33** and audio amplifier **34** for output, such as speaker output **7**, or to a headphone set, via output terminal **6** or to other remote audio output.

Decompressed display data, such as alphanumeric or animation graphic data is provided on bus **70** to display processor **36** via display memory **35** for processing and display of the image data on LCD display **2**. The display data may also be provided at video output for supply to an external TV.

The function of apparatus **1** can be expanded by the provision of additional memory employing a RAM IC card **40** (**21**A FIG. **7**) to include a software program or data for operation relative to system controller **49**. Such programs are transferred via extended interface **39** to RAM system memory **68** wherein the program can be executed under the control of system controller **49**. Also, the content on display **2** can be provided for printout to printer **41** via extended interface **39**.

Data may be received or transmitted through a modem from an external data source employing external communication interface **42** and communication cable connector **43**. Circuits **44** and **45** are relevant to the configuration shown in FIG. **6** wherein data is transmitted to antenna **20** and processed by high frequency processing circuit **45** for storage on medium **9** via digital signal processor **30** or for display via display processor **36** and audio output via processor **30**. As an example, medium **9** may contain a portion of a map for display on LCD display **2** while a signal received at antenna **20** may be superimposed on display **2** to pinpoint a particular location on the displayed map. By the same token, data on medium **9** may be read and decoded via digital signal processor **30** for transmission wherein the prepared data signal on bus **70** is modulated at circuit **44** and is transmitted from antenna **20** by high frequency signal processing circuit **45**. As an example, such a transmitted signal can indicate a particular location of apparatus **1** relative to a remote transceiver.

User interactive input to system **12** is by one or more input devices, such as touch panel **26** (FIG. **10**) or keyboard **3** or **18** (FIGS. **1**–**7**), or mouse **58** or joy stick **59**, wherein data entry is read by. system controller **49** in communication with input interface via bus **70**.

System **12** may also include, for certain additional program applications, a loading system **48** connected to system controller **49** and to disc loading unit **47** for operation and control of a microdisc, e.g., a small sized CD or CD-ROM, inserted into unit **47**.

Digital data processing unit **67** is connected to bus **70** for ECC or interleaving processing.

Lastly, power unit **52** provides for power connection to a DC power source such as a standard or rechargeable battery or an automobile battery via a DC/DC adaptor or to an external AC source via an AC adapter.

As indicated above, various different applications may be employed relative to the compact portable audio/display electronics apparatus **1** of this invention as all exemplified in FIG. **11** but provided in different possible combinations as indicated relative to the different configurations of FIGS. **1**–**7** and **10**. Also, by providing a permanently installed read/write optical medium **9** in unit **27** wherein data can be provided to the memory from an external source, such as input devices **3**, **26**, **58** and **59** or from input via IC card **40** or via antenna **46**. Lastly, the configuration of apparatus **1** can be constructed to be thin in size thereby reducing its cost and improving its structural resistance against shock and vibration.

While the invention has been described in conjunction with several specific embodiments, it is evident to those skilled in the art that many further alternatives, modifications and variations will be apparent in light of the forgoing description. Thus, the invention described herein is intended to embrace at such alternatives, modifications, applications and variations as fall within the spirit and scope of the appended claims.

What is claimed is:

**1**. An interactive, audio-visual electronic apparatus comprising:

a compact portable casing held in one hand by a user and interactively operated with the other hand of a user,

a liquid crystal display supported on a surface of said casing having a viewing surface for display of groups of matrix arranged visually displayed elements as well as pictorial display of information,

digital processing means in said casing for writing and reading audio-visual data to and from a memory in response to user invoked input via a touch key input for pictorial display of said information concurrently with reproduced audio,

said touch key input comprising a transparent matrix formed over said viewing surface of said liquid crystal display, said matrix comprising a plurality of matrix arranged transparent touch keys in corresponding alignment with said matrix arranged visually displayed elements of said liquid crystal display,

each of said touch keys functionally operating as a first set of input keys to said digital processing means in conjunction with a first set of said matrix arranged visually displayed elements and each of said touch keys functionally operating as a second, different set of input keys to said digital processing means in conjunction with a second, different set of said matrix arranged visually displayed elements after selection of at least one of said touch keys functionally operating as said first set of input keys,

a compact mass memory contained within said casing comprising an optic storage medium for recording,

storage and retrieval of audio/display digital data comprising a plurality of first data groups having a plurality of second data groups each of which is related to at least one of the first data groups, said second data groups comprising audio/visual data representations in response to an inquirable or inquisitorial query by a user through said transparent touch key input matrix relating to the plurality of first data groups, said inquirable query permitting a user relative to a particular application to inquire into a subject matter provided in said compact mass memory, said inquisitorial query permitting a user to search and examine data and interact with the displayed data via iconic representations to obtain additional displayed or audibly reproduced data,

said first set of said matrix arranged visually displayed elements being provided for selecting one of said plurality of first data groups via said digital processing means and said second set of said matrix arranged visually displayed elements being provided for selecting one of said plurality of second data groups via said digital processing means relating to the plurality of first data groups;

whereby when said touch keys are functionally operating as said second set of input keys, the second set of input keys operate in conjunction with said second set of matrix arranged visually displayed elements for allowing a user to select one of a plurality of said second data groups which comprise audio-visual representations.

2. An interactive, audio-visual electronic apparatus comprising:

a unitary, handheld casing having at least one substantially planar surface,

internally housed, removable optical storage means for storing a plurality of data groups in a random access format, each of said data groups comprising interrelated text, animation, audio and keymenu data,

a keypad for capturing user input, said keypad including a plurality of multifunctional, data specific keys programmable according to said keymenu data,

visual display means disposed on said enclosure surface for displaying at least one screen of text and animation data of a first selected data group retrieved from said optical storage means,

internally housed audio reproduction means for audibly reproducing audio data of said first selected data group synchronized to said at least one displayed screen of said first selected data group, and

an internally housed central system controller electronically coupled to said keypad, optical storage, visual display, and audio reproduction means for programming said data specific keys of said keypad in response to user input of a selection from said keymenu data of said first selected data group, retrieving a second selected data group, including interrelated text, animation, audio and keymenu data, in response to user input selection, and outputting said second selected data group data to said keypad, visual display and audio reproduction means,

wherein each of said plurality of data specific keys are on a touch sensitive matrix arranged transparent key pad and functionally operate as a first set of input keys in conjunction with said first selected data group and each of said plurality of data specific keys functionally operate as a second set of input keys in conjunction with said second selected data group.

3. A portable multimedia device, comprising:

a handheld casing housing an information processor and a mass storage unit in communication with said information processor, said mass storage unit including a medium storing multimedia data contextually arranged in a hierarchy defining a plurality of subject groups, the multimedia data comprising at least one of visual and audio data;

a display disposed on said casing and communicating with said information processor for displaying visual information to a user, the visual information including a plurality of interfacing icons;

an audio reproduction unit communicating with said information processor for reproducing audio information to the user; and

a transparent touch panel disposed on said display, said touch panel defining touch keys aligned with the interfacing icons, user activation of at least one of the touch keys causing said information processor to select one of the subject groups of the multimedia data hierarchy; and

subsequent user activation of at least one of the touch keys permits a user to inquirably and inquisitorily access multimedia data stored on the medium of said mass storage unit based on the selected one of the subject groups of the multimedia hierarchy and at least one of the visual information being displayed on said display and the audio information being reproduced by said audio reproduction unit, said inquirable query permitting a user relative to a particular application to inquire into a subject matter provided in said compact mass memory, said inquisitorial query permitting a user to search and examine data and interact with the displayed data via iconic representations to obtain additional displayed or audibly reproduced data.

* * * * *

UNITED STATES PATENT AND TRADEMARK OFFICE

# CERTIFICATE OF CORRECTION

PATENT NO. : 5,914,707
DATED : June 22, 1999
INVENTOR(S) : Mitsuru Kono

It is certified that errors appear in the above identified patent and that said Letters Patent is hereby corrected as shown below:

Title page, item 56, References Cited, U.S. Patent Documents, change "4,664,299" to --4,664,053--.

Title page, item 56, References Cited, U.S. Patent Documents, change "4,873,764" to --4,873,586--.

Signed and Sealed this

Thirtieth Day of November, 1999

*Attest:*

*Attesting Officer*

Q. TODD DICKINSON

*Acting Commissioner of Patents and Trademarks*